DUKE MATHEMATICAL JOURNAL

Volume 5, Number 2
JUNE, 1939

DUKE UNIVERSITY PRESS
DURHAM, N. C.

DUKE MATHEMATICAL JOURNAL

This periodical is published quarterly under the auspices of Duke University by Duke University Press at Durham, North Carolina. It is printed at Mt. Royal and Guilford Avenues, Baltimore, Maryland, by the Waverly Press.

Entered as second class matter at the Post Office, Durham, North Carolina. Additional entry at the Post Office, Baltimore, Maryland.

The subscription price for the current volume is four dollars, postpaid; back volumes, five dollars each, carriage extra; single numbers, one dollar and twenty-five cents, postpaid. Individual members of the Mathematical Association of America may subscribe at half price; mention of such membership should be made when subscribing. Subscriptions, orders for back numbers, and notice of change of address should be sent to Duke University Press, Durham, North Carolina.

Manuscripts and editorial correspondence should be addressed to Duke Mathematical Journal, 4785 Duke Station, Durham, North Carolina. An Author's Manual containing detailed information about the preparation of papers for publication will be sent on request.

Manuscripts submitted for consideration by the editors should be typewritten, with the exception of the formulas, which may be filled in by hand. It is recommended, though not required, that footnotes be numbered consecutively with Arabic numerals, and that they be placed at the end of the manuscript. A footnote will appear, however, at the bottom of the printed page of the journal on which reference is made to it. The author should keep a duplicate copy of the manuscript as a precaution against loss in transit and as an aid in reading proof. The original manuscript will not accompany the galley proof sent to the author unless requested. Such a request may delay the appearance of the article. Drawings must be made in black India ink on Bristol board and fully lettered.

Authors are entitled to one hundred free reprints. Additional copies will be furnished at cost. All reprints will be furnished with covers unless the contrary is specifically requested.

Made in United States of America

WAVERLY PRESS, INC.
BALTIMORE, U. S. A.

be
sic
so

ur
is
m
sp
th
a
of
cu
re
su

th
de

(2

λ(

cu

(2
or
to
p
d

ca

h

d

SURFACES OF NEGATIVE CURVATURE AND PERMANENT REGIONAL TRANSITIVITY

By Anna Grant

1. **Introduction.** The various problems connected with transitivity have been treated extensively for the flows defined by the geodesics on two-dimensional manifolds of negative curvature. A description of the extent to which solutions of the problems have been attained has been given by Hedlund [7].[1]

The manifolds in question can be obtained by identifying the points congruent under a Fuchsian group. *The present paper shows that if the Fuchsian group is of the first kind and the manifold is of negative curvature, the property of permanent regional transitivity holds.* That is, the geodesics define a flow in the space of elements such that if O is any open set of elements at time t_0, O_t is the image of O after time t, and O^* is any other open set of elements, there exists a $\bar{t}$ such that for $|t| > \bar{t}$ the set $O_t \cdot O^*$ is not empty. It is thus an extension of a similar result obtained by Hedlund [6] in the case of *constant* negative curvature. The extension requires the derivation of numerous geometric results which should be useful in the further study of the geodesic flows on the surfaces under consideration.

2. **A class of simply-connected two-dimensional manifolds.** Let U denote the unit circle $u^2 + v^2 = 1$, and let Ψ be its interior, with the following metric defined in Ψ:

$$ds^2 = \frac{\lambda^2(u, v)(du^2 + dv^2)}{(1 - u^2 - v^2)^2}, \tag{2.1}$$

$\lambda(u, v)$ of class C^m, $m \geqq 5$, and $0 < a \leqq \lambda(u, v) \leqq b$ in Ψ. The *length* of any curve segment of class C' in Ψ is $\int ds$ evaluated over the curve, ds given by (2.1). The geodesics defined by (2.1) are of at least class C^2 in arc length, coordinates of initial point, and initial direction. The term *geodesic* will refer to the geodesics defined by (2.1). Given a point in Ψ and direction at this point, there is a unique geodesic passing through the given point in the given direction.

If $\lambda(u, v) \equiv 2$ in Ψ, the geodesics are arcs of circles orthogonal to U and are called *hyperbolic lines*. Given any two points P and Q in Ψ, there is a unique hyperbolic line segment joining them; and $\int ds$ evaluated over this segment,

Received August 24, 1938; in revised form, April 8, 1939. The writer is greatly indebted to Gustav A. Hedlund for suggestions and help in the preparation of this paper.

[1] Numbers in brackets refer to the bibliography at the end of the paper.

where ds is given by (2.1) with $\lambda(u, v) \equiv 2$, is called the *hyperbolic distance* between P and Q.

A geodesic segment joining points P and Q of Ψ is *of class* A (Morse [9], p. 40) if its length is not greater than that of any other rectifiable curve joining P and Q. Given two points P and Q of Ψ, there exists a class A geodesic segment joining these points (Hedlund [5], p. 535). The *distance* between P and Q will be defined as the length of a class A geodesic segment joining P and Q. It will be denoted simply by PQ. It is easily shown that this distance satisfies the usual properties of a metric. Also, Hedlund ([5], p. 536) noted that the following fundamental theorem of Morse ([9], p. 41) holds in the case under consideration.

I. *There exists a positive constant D determined by $\lambda(u, v)$ of* (2.1) *such that in Ψ no class* A *geodesic segment can recede a hyperbolic distance greater than D from the hyperbolic line segment joining its end points.*

Any geodesic ray in Ψ can be continued to infinite length, or the geodesics in Ψ are unending. This can be proved directly, but it is a simple consequence of a result of Hopf and Rinow (cf. Hopf-Rinow [8], p. 215) which states that the above condition is equivalent to the condition that each bounded set in Ψ be compact, where the metric in terms of which bounded is defined is that already given by distance. A set in Ψ which is bounded in the sense of the hyperbolic metric is evidently compact in the sense of the hyperbolic metric. But the distance lies between two constant multiples of the hyperbolic distance (Morse [9], p. 35; Hedlund [5], p. 535), so that a set which is bounded or compact in terms of distance is bounded or compact, respectively, in terms of hyperbolic distance, and conversely. Thus a set which is bounded in terms of distance is bounded in terms of hyperbolic distance. But then it is compact in the sense of hyperbolic distance and thus compact in the sense of distance. This proves the desired result.

Let an unending geodesic be *of class* A (Morse [9], p. 40) if every finite segment of it is of class A, and let two unending curves be *of the same type* (Morse [9], p. 42) if there exists a constant δ such that any point of either is at a hyperbolic distance less than δ from some point of the other. Then the following results, essentially Morse's, hold.

II. (Morse [9], p. 44.) *Corresponding to any hyperbolic line, there exists at least one unending geodesic of class* A *of the same type. Conversely, each unending geodesic of class* A *is of the type of some hyperbolic line.*

III. (Morse [10], p. 54.) *Corresponding to each hyperbolic ray issuing from a point P in Ψ, there exists a geodesic ray of class* A *with P as initial point and of the same type as the hyperbolic ray. Conversely, each class* A *geodesic ray issuing from P is of the type of some hyperbolic ray issuing from P. There exists a constant μ, determined by $\lambda(u, v)$ of* (2.1), *such that the type distance between two class* A *geodesic rays of the same type and issuing from the same point, or between a class* A *geodesic ray and a hyperbolic ray of the same type and issuing from the same point, never exceeds μ.*

3. **The above manifold with negative curvature imposed.** Since in the metric (2.1) $E \equiv G$, $F \equiv 0$, the formula for Gaussian curvature K reduces to

$$K(u, v) = \frac{1}{2E^3}(E_u^2 + E_v^2 - EE_{uu} - EE_{vv})$$

$$= -\frac{4}{\lambda^2} + \frac{1}{\lambda^4}(1 - u^2 - v^2)^2(\lambda_u^2 + \lambda_v^2 - \lambda\lambda_{uu} - \lambda\lambda_{vv}).$$

It will be assumed that $\lambda(u, v)$ is such that

$$K(u, v) < 0, \qquad u^2 + v^2 < 1. \tag{3.1}$$

It may be noted that this condition is satisfied if $\lambda(u, v)$ is nearly constant in value.

It follows from the Gauss-Bonnet formula that the sum of the angles of a geodesic triangle is less than π and that no two geodesics can intersect twice. Given two points P and Q of Ψ, there is just one geodesic segment joining P and Q. This will be denoted by $gs(P, Q)$. The distance PQ is the length of $gs(P, Q)$ and is a continuous function of the coördinates of P and Q (cf. Bieberbach [1], p. 299). The unique geodesic on which P and Q lie will be denoted by $g(P, Q)$.

Since no two geodesics can intersect twice, all geodesics are of class A, and it follows from II, §2, that each unending geodesic g is of the type of some hyperbolic line h. The points A and B in which h meets U will be called the *points at infinity* of g. The points in Ψ at hyperbolic distance less than or equal to D from h lie in a region bounded by two circular arcs with end points A and B. It follows from I, §2, that g must lie in this region; and since g can have no multiple points, g, together with its points at infinity A and B, forms a simple continuous curve joining A and B. Thus g divides Ψ into two parts.

Similarly, from III, §2, a geodesic ray with initial point P is of the type of a hyperbolic ray with initial point P. The point in which the hyperbolic ray meets U will be called the *point at infinity* of the geodesic ray.

The geodesics through an arbitrary point P of Ψ form a field in Ψ except at P and thus geodesic polar coördinates which extend throughout Ψ can be set up with P as center. The element of distance has the form

$$\begin{cases} ds^2 = dr^2 + M^2(r, \varphi)\, d\varphi^2, & \\ M_{rr}(r, \varphi) = -K(r, \varphi)\cdot M(r, \varphi), & M_r(0, \varphi) \equiv 1, \end{cases} \tag{3.2}$$

where $K(r, \varphi)$ is the Gaussian curvature. Since $K(r, \varphi) < 0$, it follows readily that

$$M(r, \varphi) > r. \tag{3.3}$$

Let P be a point of Ψ and let g, with points at infinity A and B, be a geodesic not passing through P. Let s be the arc length on g measured from a point Q_0 and positive in the direction of B. If $Q(s)$ is the point of g with coördinate s, let $\theta(s)$ be the angle between the directed geodesic segment with initial point

$Q(s)$ and terminal point P and the geodesic ray with initial point $Q(s)$ and point at infinity A. Then the angle $\theta(s)$ is a continuous, decreasing function of s such that $\lim_{s\to-\infty} \theta(s) = \pi$ and $\lim_{s\to+\infty} \theta(s) = 0$. For assuming $\Delta s > 0$ and applying the Gauss-Bonnet formula to the geodesic triangle with vertices P, $Q(s)$ and $Q(s + \Delta s)$, we obtain

$$\theta(s) = \theta(s + \Delta s) + \alpha(s, \Delta s) - \int K\,d\omega, \tag{3.4}$$

where $\alpha(s, \Delta s)$ is the angle at P in this triangle. Since $K < 0$, $\theta(s) > \theta(s + \Delta s)$ and $\theta(s)$ is a decreasing function. Since the last two terms on the right in (3.4) approach zero with Δs, $\theta(s)$ is continuous on the right. Similarly, it can be shown that it is continuous on the left. Since $\theta(s)$ is a decreasing positive function, $\lim_{s\to+\infty} \theta(s)$ exists and let us suppose that it is not zero. With $Q(s)$ as center set up geodesic polar coördinates. For any particular value of s, the length of the geodesic segment $gs(P, Q_0)$ will equal

$$\int_0^{\theta(s)} (r'^2 + M^2\cdot\varphi'^2)^{\frac{1}{2}}\,d\varphi \geqq \int_0^{\theta(s)} M\,d\varphi,$$

and $\theta(s)$ is bounded away from zero. But by choosing s sufficiently large, it follows from (3.3) that M can be made arbitrarily large, and thus the length PQ_0 would not be bounded. We infer that $\lim_{s\to+\infty} \theta(s) = 0$. A similar argument shows that $\lim_{s\to-\infty} \theta(s) = \pi$.

If g is an arbitrary geodesic and P is an arbitrary point of Ψ, it follows from the preceding results that there exists a unique geodesic through P which meets g at right angles. It is called the geodesic through P *normal* to g, and the length of the segment of it from P to the intersection with g is the *normal distance*. From the Gauss-Bonnet formula, two geodesics which are both normal to a given geodesic g cannot intersect and thus the geodesics normal to a given geodesic form a field in Ψ. We can now set up geodesic normal coordinates using g as base, arc length along it as coördinate x, and arc length along the geodesics perpendicular to it as coördinate y. The element of distance assumes the form

$$ds^2 = M^2(x, y)\,dx^2 + dy^2. \tag{3.5}$$

If P is any point not on g, and y_P is the normal distance from P to g, any geodesic segment from P to g has length l given by

$$l = \int_{t_0}^{t_1} (\dot{y}^2 + M^2\cdot\dot{x}^2)^{\frac{1}{2}}\,dt \geqq \int_{t_0}^{t_1} (\dot{y}^2)^{\frac{1}{2}}\,dt \geqq y_P,$$

and thus the normal distance from P to g is the least geodesic distance from P to g. Let Q be the point of g such that the geodesic segment $gs(P, Q)$ is normal to g, and let s be the length on g measured from Q. It can be shown

that the distance $d(s)$ from the point P to the variable point $R(s)$ of g is a continuous function of s which is increasing for $s > 0$ and decreasing for $s < 0$ such that

$$\lim_{s\to\pm\infty} d(s) = +\infty.$$

With regard to the distance between two geodesics g and g', it is known (Hadamard [4], p. 55) that if s is the arc length measured on g, the signed normal distance $n(s)$ to g' from the point $P(s)$ of g is continuous, $-\infty < s < +\infty$, and varies according to one of the following conditions.

1. The geodesics g and g' intersect and $n(s)$ either increases from $-\infty$ to $+\infty$ or decreases from $+\infty$ to $-\infty$.

2. The geodesics g and g' do not intersect, $n(s)$ does not change sign and $|n(s)|$ either increases from 0 to $+\infty$ or decreases from $+\infty$ to 0. In this case g and g' are said to be *asymptotic*.

3. The geodesics g and g' do not intersect, $n(s)$ does not change sign, and $|n(s)|$ decreases from $+\infty$ to a positive minimum and then increases to $+\infty$.

These cases depend only on g and g' and remain the same if the rôles of g and g' in the definition of $n(s)$ are interchanged.

Since g (g') is of the type of a hyperbolic line h (h'), the normal distance between h and h' will display the same properties with respect to cases 1, 2, or 3 as does the normal distance between g and g'. Thus case 1 occurs if the points at infinity of g and g' are all distinct and separate each other on U; case 2 occurs if g and g' have just one common point at infinity; case 3 occurs if the points at infinity of g and g' are all distinct and do not separate each other on U. In particular we conclude that g and g' cannot have the same points at infinity and therefore there is just one geodesic of the type of a given hyperbolic line. The unique geodesic with points at infinity A and B will be denoted by $g(A, B)$.

It follows from I and III, §2, that two geodesic rays issuing from the same point P of Ψ must have distinct points at infinity. The unique geodesic ray with initial point P and point at infinity A will be denoted by $gr(P, A)$.

Given A on U and P in Ψ, there exists a geodesic passing through P and having A as one of its points at infinity. Since no two geodesics with A as point at infinity can intersect in Ψ, the geodesics with a common point at infinity form a field in Ψ.

The distance between two points of Ψ has been defined. A sequence of points in Ψ approaches a point P of Ψ if the distance from P of the points of the set approaches zero. The distance between two points of U will be defined as Euclidean distance, and a point varies continuously on U if its distance from a fixed point of U varies continuously. A sequence of points within or on U shall be said to approach a point A of U as limit point if the Euclidean distance of the points of the sequence from A approaches zero.

An *element* is a point of Ψ, $P(u, v)$, and a direction φ at this point, where it can be assumed that φ is measured from the direction parallel to the positive u-axis. An element thus has three coördinates and will be denoted by $p(u, v, \varphi)$.

The point $P(u, v)$ will be said to be the point *bearing* $p(u, v, \varphi)$. The *distance* between the elements $p(u_1, v_1, \varphi_1)$ and $q(u_2, v_2, \varphi_2)$ will be defined as the sum of the distance between (u_1, v_1) and (u_2, v_2) and $\| \varphi_2 - \varphi_1 \|$ where this denotes the least numerical value of the set $| \varphi_2 - \varphi_1 + 2n\pi |$ $(n = 0, \pm 1, \cdots)$. Continuity and limit point in the space of elements are understood to be defined in terms of this metric.

The element $e(u, v, \varphi)$ determines a unique geodesic ray, namely, that one with initial element e. If A is the point at infinity of this ray, it can be shown that, as a consequence of III, §2, and the continuous variation of a geodesic with continuous variation of the initial conditions, A varies continuously on U as e varies continuously. Also, if P and A vary continuously, the direction of $gr(P, A)$ at P varies continuously, and if Q in Ψ approaches A on U, the direction of $gs(P, Q)$ at P approaches that of $gr(P, A)$ at P.

Let g be a geodesic with points at infinity A and B and let C be a point of U distinct from A and B. Let s be the arc length on g measured from a point Q_0 on g and positive in the direction of B. If $Q(s)$ is the point of g with coördinate s, let $\theta(s)$ be the angle at $Q(s)$ between the directed geodesic rays $gr(Q(s), C)$ and $gr(Q(s), A)$. Then $\theta(s)$ is a continuous monotonic decreasing function of s such that $\lim_{s \to -\infty} \theta(s) = \pi$, $\lim_{s \to +\infty} \theta(s) = 0$. For by choosing the point P sufficiently close to C, the angle $\varphi(s)$ at $Q(s) = Q$ between $gs(Q, P)$ and $gr(Q, A)$ can be made arbitrarily close to $\theta(s)$, and the angle $\varphi(s + \Delta s)$ at $Q(s + \Delta s) = Q'$ between $gs(Q', P)$ and $gr(Q', A)$ can be made arbitrarily close to $\theta(s + \Delta s)$. But it has been shown that $\varphi(s)$ is a decreasing function of s and it follows that $\theta(s)$ must be monotonic decreasing. That $\theta(s)$ is continuous follows from the fact that as Δs approaches zero, $Q(s + \Delta s)$ approaches $Q(s)$, and the initial element of $gr(Q(s + \Delta s), C)$ approaches the initial element of $gr(Q(s), C)$. To show that $\lim_{s \to +\infty} \theta(s) = 0$, let R be a fixed point of $g(B, C)$. Then $\theta(s)$ is less than the angle $\alpha(s) = RQ(s)A$. As s becomes positively infinite, $\alpha(s)$ approaches zero and consequently $\theta(s)$ must approach zero. A similar argument shows that $\lim_{s \to -\infty} \theta(s) = \pi$.

4. **Geodesic circles.** If geodesic polar coördinates are set up with any point Q of Ψ as base, the curve $r =$ constant will be called a *geodesic circle* with center Q. For brevity the term *circle* will mean geodesic circle except in the case of the unit circle U. A circle is a simple closed curve of class C^2 and is perpendicular at a point of it to the geodesic determined by the point and the center of the circle.

THEOREM 4.1. *A circle and a geodesic intersect in two points, or are tangent and have one point in common, or have no points in common.*

This is an immediate consequence of the way in which the distance from a fixed point to a point of a geodesic varies.

A domain Φ of Ψ is *convex* if, P and Q being arbitrary points of Φ, the geodesic segment $gs(P, Q)$ lies entirely in Φ.

THEOREM 4.2. *A circle bounds a convex domain.*

For if P and Q are interior points of the circle C, and A and B are points at infinity of $g(P, Q)$ such that the order of points on $g(P, Q)$ is $APQB$, then $gr(P, A)$ and $gr(Q, B)$ must both intersect C. It follows from Theorem 4.1 that the segment $gs(P, Q)$ can have no point on C and thus all of its points must be interior to C.

THEOREM 4.3. *If P and Q are points of the circle C, the points of $gs(P, Q)$ other than P and Q are all interior points of C, the points of $g(P, Q)$ not on $gs(P, Q)$ are all exterior points of C.*

For the points of $gs(P, Q)$ other than P and Q are all nearer the center of C than either P or Q. The points of $g(P, Q)$ not on $gs(P, Q)$ are all farther away from the center of C than either P or Q.

THEOREM 4.4. *Two circles have no points in common, or are tangent and have one common point, or intersect in two points. In the last case the two points of intersection lie on opposite sides of the geodesic determined by the centers.*

Two circles with the same center and different radii have no common point. If two circles are tangent at P, it is readily shown that, except for P, C' lies either entirely exterior or entirely interior to C. That two circles which have only one common point must be tangent is clear from the fact that if they are not tangent, each must have on it points interior and exterior to the other. Thus if C and C' are the two circles and P and Q are points of C, respectively interior and exterior to C', the two arcs of C determined by P and Q must intersect C' in distinct points.

It remains to show that two circles can intersect in at most two points. Suppose that C and C' intersect in three points, P_1, P_2, and P_3. The centers Q and Q' of C and C', respectively, cannot coincide. No one of the points P_1, P_2, P_3 can lie on $g(Q, Q')$; for, if it did, the triangle property would exclude the existence of any other common point. Therefore two of them, say P_1 and P_2, lie on the same side of $g(Q, Q')$ but not on it. Since P_1 and P_2 are assumed distinct, $g(Q, P_1)$ and $g(Q, P_2)$ can have only Q in common. Similarly, $g(Q', P_1)$ and $g(Q', P_2)$ can have only Q' in common. It can be assumed that the notation is so chosen that P_1 is exterior to the geodesic triangle $QQ'P_2$. Then $gs(Q, P_2)$ cannot cut across $gs(Q', P_1)$; for, if there were such a point of intersection R, it would follow that

$$P_2R > | Q'P_2 - Q'R | = | Q'P_1 - Q'R | = P_1R,$$

as well as

$$P_1R > | QP_1 - QR | = | QP_2 - QR | = P_2R,$$

and both of these cannot hold. Therefore the geodesic triangle QP_1Q', which constitutes a simple Jordan curve, would contain P_2 in its interior and $g(P_1, P_2)$, which cannot intersect $g(Q, P_1)$ or $g(Q', P_1)$ except at P_1, would necessarily cut $gs(Q, Q')$ in some point S. Since $QQ' < QP_1 + Q'P_1$, the sum of the radii of C and C', S would necessarily lie within one of the circles and Theorem 4.3 would be contradicted.

The proof of the theorem is complete.

THEOREM 4.5. *If two circles intersect in one or two points, the geodesic segment determined by their centers can have no point exterior to both circles.*

If the circles C and C' have centers Q and Q' and radii r and r', respectively, the condition that they have a common point implies that $r + r' \geqq QQ'$. But if S is any point of $gs(Q, Q')$, $QS + Q'S = QQ' \leqq r + r'$, and either $QS \leqq r$ or $Q'S \leqq r'$. This implies the stated result.

5. **Equidistant curves.** In the case of constant negative curvature the locus of points equidistant from two points P and Q of Ψ is the hyperbolic line which bisects perpendicularly $gs(P, Q)$. In the case of variable negative curvature this locus is not necessarily a geodesic, but it can be shown that it has some of the properties of a geodesic as far as behavior in the large is concerned.

THEOREM 5.1. *The locus of points equidistant from two given points of Ψ is a continuous unending curve which is the topological image of a line and is of the type of a hyperbolic line.*

Let Q and Q' be the two given points and let $2d = QQ'$. If P is an arbitrary point of Ψ, let r denote its distance from Q, r' its distance from Q'. Then every point of Ψ has a coördinate pair (r, r'). If $r < d$, the geodesic circles with centers Q and Q' and radius r have no point in common, and thus no point of Ψ has coördinates (r, r) if $r < d$. The midpoint of $gs(Q, Q')$ has the coördinates (d, d), and it is evidently the only point of Ψ with these coördinates. If $r > d$, the circle C_r with center Q and radius r intersects $g(Q, Q')$ in two points of which one is at distance $| r - 2d |$ and the other at distance $r + 2d$ from Q'. Thus there are points of C_r interior and exterior to the circle C'_r with center Q' and radius r, and C_r intersects C'_r. It follows from Theorem 4.4 that C_r intersects C'_r in just two points and these are on opposite sides of $g(Q, Q')$. Let σ denote the closed region bounded by $g(Q, Q')$ and one of the arcs of U determined by the points at infinity of $g(Q, Q')$. Then corresponding to $r \geqq d$, there is just one point of σ with the coördinates (r, r) and of these only (d, d) is on the boundary of σ. It is sufficient to show that the set $R(r, r)$, $r \geqq d$, of σ is the topological image of a ray and is of the type of a hyperbolic ray.

It has been shown already that the set R is in one-to-one correspondence with the set of values of r, $r \geqq d$. To show that this correspondence is continuous, let $P(r, r)$ and $P'(r + \Delta r, r + \Delta r)$ be points of σ. If we did not have $\lim_{\Delta r \to 0} P' = P$, because of the compactness of the space there would exist a point P^*,

different from P, in σ and with coördinates (r, r). But this is impossible. Conversely, nearby points in Ψ have coördinate pairs differing only slightly and the correspondence between the set R and the set $r \geqq d$ is topological.

It remains to show that the set R is of the type of a hyperbolic ray. To that end, let T with coördinates (t, t), $t > d$, be a point of σ, let $g(Q, Q')$ have points at infinity A and A' such that the order on $g(Q, Q')$ is $AQQ'A'$, and let B_t and B_t' be the points at infinity of $gr(Q, T)$ and $gr(Q', T)$, respectively. The geodesic $g(Q, Q')$ and the rays $gr(Q, B_t)$ and $gr(Q', B_t')$ divide the part of Ψ in σ into four parts which can conveniently be denoted by the vertices on their boundaries. If X is any point, other than T, of $A'Q'TB_t$, $gs(Q, X)$ must cut $gs(Q', T)$ in some point Y and

$$Q'Y = Q'T - YT = QT - YT < QY.$$

Thus

$$Q'X < Q'Y + YX < QY + YX = QX,$$

and no point, other than T, of $A'Q'TB_t$ can be a point of R. A similar statement holds for $AQTB_t'$.

If X is any point, other than T, of $QQ'T$, $gr(Q, X)$ must cut $gs(Q', T)$ in some point Z and

$$QX + Q'X < QX + XZ + Q'Z = QZ + Q'Z < QT + ZT + Q'Z$$
$$= QT + Q'T = 2QT,$$

and at least one of the lengths QX and $Q'X$ must be less than $QT = t$. It follows that the points (r, r) of R which lie in $QQ'T$ must have $r < t$. If we combine these results, the points (r, r) of R for which $r > t$ must lie in the interior of $TB_t'B_t$. As t increases, the points B_t and B_t' move toward each other on U and each must approach a limiting point. If $\lim_{t\to\infty} B_t = B$, then $\lim_{t\to\infty} T = B$, and since T lies on $gr(Q', B_t')$, it follows that $\lim_{t\to\infty} B_t' = B$.

The set R, except for (d, d), lies in the interior of the region bounded by $gs(Q, Q')$ and the geodesic rays $gr(Q, B)$ and $gr(Q', B)$. For all points of R lie in the region bounded by $gs(Q, Q')$, $gr(Q, B_t')$, $gr(Q', B_t)$, and the arc $B_t'B_t$ of U, no matter what the value of t. Since as t becomes infinite $gr(Q, B_t')$ approaches $gr(Q, B)$ and $gr(Q', B_t)$ approaches $gr(Q', B)$, the set R must lie either in or on the boundary of the stated region. The point (d, d) is the only point of R on $gs(Q, Q')$. Suppose there were a point $T(t, t)$ of R on either $gr(Q, B)$ or $gr(Q', B)$. This would imply that either B_t or B_t' must coincide with B. But as t increases, both of the points B_t and B_t' move, and neither can coincide with B. The set R satisfies the stated condition.

It follows that the set R is of the type of $gr(Q, B)$, or equally well $gr(Q', B)$, and the proof of the theorem is complete.

The locus of points equidistant from the points P and Q of Ψ will be denoted

by $E(P, Q)$ and the points at infinity of the hyperbolic line which is of the type of $E(P, Q)$ will be called the *points at infinity* of $E(P, Q)$.

The following theorem will be useful.

THEOREM 5.2. *Let Q and Q' be points of Ψ and D a point of U not identical with either of the points at infinity of $E(Q, Q')$. Then there exists a neighborhood of D and a positive constant δ such that if P is any point of Ψ in this neighborhood,*

$$|PQ - PQ'| > \delta.$$

Let A and A' be the points at infinity of $g(Q, Q')$ such that the order of points is $AQQ'A'$. If D coincides with A', a neighborhood of D can be chosen so that for every P of Ψ in this neighborhood, $gs(Q, P)$ has on it a point R within distance $\frac{1}{3}d$ of Q'. But then

$$PQ - PQ' = PR + RQ - PQ' > \tfrac{2}{3}d - \tfrac{1}{3}d = \tfrac{1}{3}d.$$

A similar proof applies if D coincides with A.

If we assume that D is neither A nor A', D lies in the interior of one of the four intervals of U determined by A, A', and the points at infinity B, B', of $E(Q, Q')$. The proofs are similar in the four cases and it can be assumed that D lies in $A'B$. Then $gr(Q, D)$ cuts across $gr(Q', B)$ in some point M. If the closed subinterval ζ of $A'B$ with midpoint D is chosen sufficiently small, all the geodesic rays with initial point Q and point at infinity in ζ will intersect $gr(Q', B)$ in points of a closed interval γ. No point of R is in γ, and by use of the method of proof of Theorem 5.1 it is readily shown that if T is a point of γ, $QT > Q'T$. Since γ is closed and $QT - Q'T$ varies continuously as T varies continuously, there exists a $\delta > 0$ such that for T in γ, $QT - Q'T > \delta$. Let the neighborhood of D be chosen so that if P is a point of Ψ in it, P lies in the region bounded by ζ and the geodesic with points at infinity at the ends of ζ. For any such P, $gs(Q, P)$ cuts across γ in a point T and

$$QP - Q'P = QT + TP - Q'P > QT - Q'T > \delta.$$

The stated theorem is proved.

6. Horocycles. Existence and some geometric properties.[2]

In the case of constant negative curvature, a horocycle is a Euclidean circle tangent to U at some point A and is an orthogonal trajectory of the field of hyperbolic lines with A as point at infinity. The first of these characterizations evidently cannot be extended to the case of variable negative curvature and the second is not a convenient point of departure, though the geodesics with A as one point at infinity do form a field in Ψ. However, in the case of constant negative curvature, a horocycle is the limit of hyperbolic circles passing through a fixed point of Ψ as the centers approach a point A of U. The analogue of this will

[2] The author is indebted to the referee for the proofs of the lemmas and theorems of this section. The original proofs did not make use of convex sets and were considerably longer.

be used to define the generalized horocycles which will be found to possess many of the properties of the (hyperbolic) horocycles.

Let $N_r(P)$ be the r-neighborhood of P, that is, the totality of points at distance less than r from P. The boundary of $N_r(P)$ is a circle. For any P of Ψ and A of U, let $P(s)$ denote the point of $gr(P, A)$ at distance s ($s \geqq 0$) from P. Let the *horocyclic region* $\bar{C}(P, A)$ be defined (cf. Busemann [3], pp. 144–145) as the point set sum $\sum_{s>0} N_s(P(s))$.

THEOREM 6.1. *The set $\bar{C}(P, A)$ is an open convex set.*

Since $\bar{C}(P, A)$ is the sum of open sets, it is open. If Q_1 and Q_2 are any two points of $\bar{C}(P, A)$, it follows from the definition of $\bar{C}(P, A)$ that there exist positive numbers s_1 and s_2 such that Q_i is in $N_{s_i}(P(s_i))$ ($i = 1, 2$). With the aid of the triangle inequality we can easily show that if $s_1 \geqq s_2$, then $N_{s_1}(P(s_1)) \supset N_{s_2}(P(s_2))$, and hence if s is the larger of s_1 and s_2, both points Q_1 and Q_2 are in $N_s(P(s))$. But $N_s(P(s))$, $s > 0$, lies in $\bar{C}(P, A)$ and is convex. This completes the proof of the theorem.

LEMMA 6.1. *If $gr(P, A)$ and $gr(P, B)$ are orthogonal at P, and $\theta(s)$ is the acute angle between $gr(P(s), B)$ and $gs(P(s), P)$ ($s > 0$), then for every point Q on $gr(P, B)$*

$$QP(s_2) - QP(s_1) \geqq \int_{s_1}^{s_2} \cos \theta(s)\, ds > 0,$$

whenever $0 \leqq s_1 < s_2$.

For any point Q on $gr(P, B)$ let $\phi(s)$ be the angle between $gs(P(s), Q)$ and $gs(P(s), P)$. As Q tends to B along $gr(P, B)$, $\phi(s)$ increases and approaches $\theta(s) < \frac{1}{2}\pi$. It follows that (cf. Bliss [2], p. 100)

$$\frac{d}{ds} QP(s) = \cos \phi(s) > \cos \theta(s), \qquad s > 0,$$

whence by integration the lemma holds for $0 < s_1 < s_2$. By continuity it holds if $s_1 = 0$.

THEOREM 6.2. *In geodesic coördinates (x, y) with $gr(P, A)$ as positive x-axis, the horocyclic region $\bar{C}(P, A)$ is defined by inequalities $y^-(x) < y < y^+(x)$, $x > 0$, and the boundary of $\bar{C}(P, A)$ consists of the two arcs $y = y^+(x)$ and $y = y^-(x)$, where $y^+(x)$ and $y^-(x)$ are continuous functions defined for $x \geqq 0$ and $y^+(0) = y^-(0) = 0$.*

All points $(\bar{x}, y)$, $\bar{x} < 0$, are exterior points of $N_x(P(x)) = N_x(x, 0)$, $x > 0$, so that any such point is not a point of the set $\bar{C}(P, A)$. All points $P(x)$, $x > 0$, are in $\bar{C}(P, A)$, but P is not, so P is on the boundary of $\bar{C}(P, A)$ which will be denoted by $C(P, A)$. If the point $(0, y)$, $y \neq 0$, is at distance d_{xy} from $(x, 0)$, it follows from Lemma 6.1 that there exists a positive constant $\delta(y)$ such that $d_{xy} - x \geqq \delta(y) > 0$. Hence the $\delta(y)$-neighborhood of $(0, y)$ has no

points in $N_x(P(x))$ and the distance from $(0, y)$ to the set $\bar{C}(P, A)$ is at least $\delta(y)$. Thus, for every point of $\bar{C}(P, A)$, $x > 0$, and the only point of $C(P, A)$ with $x = 0$ is P itself.

Choose any $x_0 > 0$ and let B and C be the points at infinity of the geodesic $x = x_0$. The point $(x_0, 0)$ is in $\bar{C}(P, A)$ and since $\bar{C}(P, A)$ is convex, it follows that the part of $x = x_0$ in $\bar{C}(P, A)$ is an open segment or an open ray or an entire geodesic. Let X_0 be the point $(x_0, 0)$ and let $Y(y)$, $y > 0$, denote the point (x_0, y), where we assume the notation has been so chosen that $Y(y)$ approaches B as y becomes positively infinite. If $\theta(y)$ denotes the acute angle between $gs(Y(y), X_0)$ and $gr(Y(y), A)$, $\lim_{y \to +\infty} \theta(y) = 0$ and there exists a $y_0 > 0$ such that

$$\int_0^{y_0} \cos \theta(y)\, dy > PX_0 + 1.$$

Now if Q is any point of $gr(X_0, A)$, it follows from Lemma 6.1 that

$$QY(y_0) \geqq QX_0 + \int_0^{y_0} \cos \theta(y)\, dy \geqq QP + 1.$$

Thus the distance from $Y(y_0)$ to the set $\bar{C}(P, A)$ is at least 1 and not all of $gr(X_0, B)$ is in $\bar{C}(P, A)$. Similarly, not all of $gr(X_0, C)$ is in $\bar{C}(P, A)$ and the part of $g(B, C)$ in $\bar{C}(P, A)$ is a finite open segment which may be denoted by $y^-(x_0) < y < y^+(x_0)$.

It is easily shown that the convexity of the set $\bar{C}(P, A)$ and the finiteness of $y^-(x)$ and $y^+(x)$ for each $x \geqq 0$ imply the continuity of these functions. This completes the proof of the theorem.

The set $C(P, A)$ of points $(x, y^+(x))$ and $(x, y^-(x))$ will be called the *horocycle determined by P and A*. As $\bar{x}$ becomes positively infinite, the points at infinity of the geodesic $x = \bar{x}$ both move toward A. They must both approach A as limit point for, if either one did not, $x = \bar{x}$ could not remain perpendicular to $gr(P, A)$. It follows that the whole of the geodesic $x = \bar{x}$ shrinks to A as $\bar{x}$ becomes infinite and consequently $\lim_{x \to +\infty} (x, y^+(x)) = A$ and $\lim_{x \to +\infty} (x, y^-(x)) = A$. The point A will be called the *point at infinity* of $C(P, A)$. The horocycle $C(P, A)$ together with A forms a simple closed curve of which $\bar{C}(P, A)$ is the interior.

The set $N_r(P)$, $r > 0$, has been defined as the set of points at distance less than r from P. With the understanding that $N_r(P)$ is empty if $r \leqq 0$, we can state the following lemma.

LEMMA 6.2. *If $\rho(s)$ is a non-negative function defined for $s > 0$ and if $\lim_{s \to +\infty} \rho(s) = 0$, then*

$$\sum_{s>0} N_{s-\rho(s)}(P(s)) = \bar{C}(P, A).$$

Since $N_{s-\rho(s)}(P(s)) \subset N_s(P(s)) \subset \bar{C}(P, A)$, it follows that $\sum_{s>0} N_{s-\rho(s)}(P(s)) \subset \bar{C}(P, A)$. Conversely, let Q be any point of $\bar{C}(P, A)$. From the definition of

$\bar{C}(P, A)$ we can infer the existence of $s_0 > 0$ such that $P \subset N_{s_0}(P(s_0))$, and since this is an open set, there exists a $\delta > 0$ such that $N_\delta(Q) \subset N_{s_0}(P(s_0))$. If s is so chosen that $s > s_0$ and $\rho(s) < \delta$, we have $N_s(P(s)) \supset N_{s_0}(P(s_0)) \supset N_\delta(Q)$ and hence Q lies in $N_{s-\delta}(P(s)) \subset N_{s-\rho(s)}(P(s)) \subset \sum_{s>0} N_{s-\rho(s)}(P(s))$. Thus $\bar{C}(P, A) \subset \sum_{s>0} N_{s-\rho(s)}(P(s))$, and the proof is complete.

LEMMA 6.3. *If Q is on $C(P, A)$, then $C(Q, A) = C(P, A)$.*

Let $P(s)$ be the point on $gr(P, A)$ with $PP(s) = s$, and let $Q(\sigma)$ be the point on $gr(Q, A)$ with $QQ(\sigma) = \sigma$. It will first be shown that $\rho(s) = P(s)Q(s) \to 0$ as $s \to \infty$.

Since the two geodesic rays are asymptotic, for every $\delta > 0$ there is an $s(\delta)$ such that if $s \geqq s(\delta)$ there is a point $Q'(s)$ on $gr(Q, A)$ with $P(s)Q'(s) < \delta$. Also, since Q is on $C(P, A)$, $s(\delta)$ can be taken so large that for all $s \geqq s(\delta)$, $0 \leqq QP(s) - s < \delta$. But $| QQ'(s) - QP(s) | < P(s)Q'(s) < \delta$, so $| s - QQ'(s) | < 2\delta$. By definition of $Q(s)$, this gives $Q(s)Q'(s) < 2\delta$; hence $\rho(s) = P(s)Q(s) < 3\delta$, and $\rho(s) \to 0$ as $s \to \infty$.

By the definition of $\rho(s)$, $N_{s-\rho(s)}(P(s)) \subset N_s(Q(s)) \subset \bar{C}(Q, A)$. Using Lemma 6.2, we see that $\bar{C}(P, A) = \sum_s N_{s-\rho(s)}(P(s)) \subset \bar{C}(Q, A)$. If we interchange the rôles of P and Q, $\bar{C}(Q, A) \subset \bar{C}(P, A)$. Hence $\bar{C}(P, A) = \bar{C}(Q, A)$, and the lemma is established.

THEOREM 6.3. *The horocycle $C(P, A)$ has a continuously turning tangent at every finite point and is an orthogonal trajectory of the geodesics with A as point at infinity.*

From Lemma 6.3 we see that it is sufficient to prove the theorem for the point P.

On account of the convexity of $\bar{C}(P, A)$, as $x \to 0$, $g(P, (x, y^+(x)))$ must rotate in the counterclockwise sense and must therefore approach a limiting geodesic through P. Similarly, the geodesics determined by P and the points of $y^-(x)$ approach a limiting geodesic through P. Since $C(P, A)$ is the limit of circles through P perpendicular to $g(P, A)$, these limits are the same and coincide with the geodesic through P perpendicular to $g(P, A)$. Thus $C(P, A)$ has a tangent direction at each point and is an orthogonal trajectory of the field of geodesics with A as point at infinity.

That the curve has a continuously turning tangent is now an immediate consequence of the continuous variation of the geodesics determined by A and points which approach P along the continuous curve $C(P, A)$.

7. **Approximation properties of the horocycles.** A horocycle $C(P, A)$ is determined by a point P of Ψ and a point A of U. In this section it will be shown that a point of $C(P, A)$ is approximated arbitrarily closely by points of horocycles determined by P' near P and A' near A. The following theorem is easily proved (cf. Busemann [3], p. 145).

THEOREM 7.1. *A point Q is on $C(P, A)$ if and only if it is the limit point of points on circles with centers on $gr(P, A)$ and passing through P, as the radii of the circles become infinite.*

THEOREM 7.2. *Given any point Q of the horocycle $C(P, A)$ and $\epsilon > 0$, there exists a $\delta > 0$ such that every circle with center within distance δ of A and having on it a point within distance δ of P has on it a point within distance ϵ of Q. Conversely, if Q is a point of Ψ and a limit point of points on a sequence of circles C_n with centers approaching A, and if C_n has on it a point P_n such that $\lim_{n \to \infty} P_n = P$, Q is on $C(P, A)$.*

It will first be shown that given $\epsilon > 0$, there exists a $\delta > 0$ such that if R is any point of Ψ at distance less than δ from A, $| RP - RQ | < \frac{1}{2}\epsilon$. For it is clear from Theorem 5.2 that A is a point at infinity of $E(P, Q)$, therefore $E(P, Q)$ has on it points arbitrarily near A and between $gr(P, A)$ and $gr(Q, A)$ thus arbitrarily close to both of these rays. Let $M \subset E(P, Q)$ and be such that $gr(P, A)$ and $gr(Q, A)$ both cross an $\frac{1}{8}\epsilon$-neighborhood of M. Then $\delta > 0$ can be chosen so small that if $RA < \delta$, $gs(P, R)$ and $gs(Q, R)$ will have points S and T, respectively, in the $\frac{1}{8}\epsilon$-neighborhood of M. But then

$$| RP - RQ | = | RS + SP - RT - TQ |$$
$$\leqq | RS - RT | + | SP - MP | + | MQ - TQ | < \tfrac{1}{2}\epsilon.$$

Now let $\delta > 0$ also be chosen less than $\frac{1}{2}\epsilon$. Let C be a circle with center R within distance δ of A and having on it a point N within distance δ of P. Then we have

$$| RN - RQ | \leqq | RN - RP | + | RP - RQ | < \delta + \tfrac{1}{2}\epsilon + \epsilon.$$

The first statement of the theorem is proved.

To prove the remainder of the theorem, suppose that the stated conditions are satisfied and Q is not on $C(P, A)$. Then A cannot be a point at infinity of $E(P, Q)$; for if it were, there would be points arbitrarily close to A and between $gr(P, A)$ and $gr(Q, A)$, thus arbitrarily close to $gr(P, A)$, and at the same distance from P and Q. But then Q would be a limit point of points on circles with centers on $gr(P, A)$ and with radii arbitrarily large, and according to Theorem 7.1 would be on $C(P, A)$.

If A is not a point at infinity of $E(P, Q)$, there exist according to Theorem 5.2 a $\delta > 0$ and a neighborhood of A such that if R is any point in Ψ and in this neighborhood, $| RP - RQ | > \delta$. But then $| RP_n - RQ | > \delta - | RP_n - RP |$ and there exists an N such that

$$| RP_n - RQ | > \tfrac{1}{2}\delta > 0, \qquad n > N.$$

The point Q cannot be a limit point of the stated kind, and the proof of the theorem is complete.

THEOREM 7.3. *Given any point Q of the horocycle $C(P, A)$ and $\epsilon > 0$, there exists a $\delta > 0$ such that every horocycle with point at infinity within distance δ of A and having on it a point within distance δ of P has on it a point within distance ϵ of Q. Conversely, if Q is a limit point of points on a sequence of horocycles $C(P_n, A_n)$ such that $\lim_{n\to\infty} P_n = P$ and $\lim_{n\to\infty} A_n = A$, Q is on $C(P, A)$.*

Let Q be a point of $C(P, A)$. Given $\epsilon > 0$, let $\delta > 0$ be chosen in accordance with Theorem 7.2. Let $C(M, B)$ be a horocycle such that $MP < \delta$ and $AB < \delta$. Then if R is on $gr(M, B)$, for all MR exceeding some constant, $RA < \delta$. According to Theorem 7.2 every circle with center at such a point R and passing through M has on it a point S_r within distance ϵ of Q. As R approaches B, the set S_r must have a point of accumulation S within distance ϵ of Q, and it follows from Theorem 7.1 that S is a point of $C(M, B)$. The first part of the theorem is proved.

If Q is a limit point of points on a sequence of horocycles $C(P_n, A_n)$ such that $\lim_{n\to\infty} P_n = P$ and $\lim_{n\to\infty} A_n = A$, by making use of the first part of Theorem 7.1, we infer that Q is a limit point of points on a sequence of circles C_n with centers approaching A and such that P_n is on C_n. The stated result then follows from Theorem 7.2.

8. Further geometric properties of the horocycles.

THEOREM 8.1. *Given two points P and Q of Ψ, there are just two horocycles containing both P and Q. The points at infinity of these horocycles are the points at infinity of $E(P, Q)$.*

Let A and B be the points at infinity of $E(P, Q)$ and let M_r be the point of $gr(P, A)$ at distance r from P and C_r the circle with center M_r and radius r. Then P is on C_r. Since $E(P, Q)$ lies in the region bounded by $gr(P, B)$, $gr(Q, B)$, $gr(P, A)$, and $gr(Q, A)$, given $\epsilon > 0$, there exists an $\bar{r}$ such that if $r > \bar{r}$, there is a point N_r, on $E(P, Q)$ and within distance ϵ of M_r. Then

$$|M_rQ - M_rP| \leqq |M_rQ - N_rQ| + |N_rQ - N_rP| + |N_rP - M_rP| < 2\epsilon.$$

The point Q is a limit point of points of circles C_r with arbitrarily large radius. From Theorem 7.1, Q must lie on $C(P, A)$.

A similar proof applies to $C(P, B)$ and the existence of two horocycles with the stated properties is proved.

To complete the proof it is sufficient to show that any horocycle which contains P and has neither A nor B as its point at infinity cannot contain Q. Let such a horocycle be $C(P, D)$. According to Theorem 5.2, there exist a neighborhood of D and a $\delta > 0$ such that if R is any point of Ψ in this neighborhood, $|RP - RQ| > \delta$. But then Q cannot be a limit point of points on circles through P, with centers on $gr(P, D)$ and arbitrarily large radii, and consequently, from Theorem 7.1, Q cannot be on $C(P, D)$.

THEOREM 8.2. *There is one and only one point of a horocycle with A as point at infinity on each geodesic with A as point at infinity.*

Let P be any point of the given horocycle and let $g(C, A)$ be any geodesic with A as point at infinity. Since $g(C, A)$ and $g(P, A)$ are asymptotic, there are on $g(C, A)$ points which are interior to $C(P, A)$. But A is the only point of U on $C(P, A)$, therefore $C(P, A)$ must cut $g(C, A)$ in a finite point.

If there were two such points, say Q and Q', then $C(Q, A) = C(P, A) = C(Q', A)$, and this is impossible. The stated result is proved.

The following theorem is implied by more general considerations of Busemann ([3], pp. 145–146).

THEOREM 8.3. *Two horocycles with the same point at infinity A cut off equal intercepts on the geodesics with point at infinity A.*

THEOREM 8.4. *The finite points of intersection of a geodesic and a horocycle fulfill one, and only one, of the following conditions:*

(a) *there are none;*

(b) *there is one, in which case the geodesic and horocycle are either tangent at the common point, or orthogonal at the common point and have the same point at infinity;*

(c) *there are two, in which case the geodesic and horocycle are neither tangent nor orthogonal.*

Let the points at infinity of the geodesic be B and C and let the horocycle be $C(P, A)$. If A coincides with either B or C, it follows from Theorem 8.2 that $g(B, C)$ and $C(P, A)$ have just one point in common and from Theorem 6.3 that they are orthogonal at this point. This is one of the possibilities under (b).

Assuming that A does not coincide with either B or C, we see that there exists a point D of U such that $g(A, D)$ and $g(B, C)$ are orthogonal. The horocycle $C(P, A)$ and the geodesic $g(A, D)$ have just one point Q in common, and if we make Q the origin and $gr(Q, A)$ the positive x-axis in a geodesic normal coördinate system, $C(P, A)$ is given by functions of class C', $y = y^{+}(x)$ and $y = y^{-}(x)$, $x \geqq 0$, $y^{+}(0) = y^{-}(0) = 0$. But the coördinate system has been so chosen that $g(B, C)$ has the equation $x = \bar{x}$. It follows that if $\bar{x} < 0$, $g(B, C)$ and $C(P, A)$ have no point in common. If $\bar{x} = 0$, $g(B, C)$ and $C(P, A)$ have just one point Q in common, and since $g(B, C)$ is orthogonal to $gr(Q, A)$ and the same is true of $C(P, A)$, $g(B, C)$ and $C(P, A)$ must be tangent. Finally, if $\bar{x} > 0$, $y^{-}(\bar{x}) < 0 < y^{+}(\bar{x})$ and $g(B, C)$ and $C(P, A)$ have just two points in common. At neither of these points can $g(B, C)$ and $C(P, A)$ be normal, for this would imply the coincidence of A with either B or C. At neither of these points can they be tangent, for then one of the functions $y^{-}(x)$ or $y^{+}(x)$ would not have a derivative at $x = \bar{x}$.

9. **Element approximation.** Since a horocycle is a curve of class C', if oriented it bears an element at each of its points. Various theorems concerning element approximation are needed for the derivation of the permanent regional transitivity, and these will be treated now.

A circle is a *right* circle if it is so oriented that if P is a point of it and R is its center, the relation of the initial direction of the directed segment $gs(P, R)$ to the direction of the circle at P is the same as that of the positive x-axis to the positive y-axis. The circle with the opposite orientation is a *left* circle.

A horocycle is a *right* horocycle if it is so oriented that if P is a point of it and A is its point at infinity, the relation of the initial direction of the directed geodesic ray $gr(P, A)$ to the direction of the horocycle at P is the same as that of the positive x-axis to the positive y-axis. The notation will be $C_R(P, A)$. A *left* horocycle has the opposite orientation and the notation $C_L(P, A)$.

It is to be understood in the following, and indeed it was implicit before, that whenever the notations $gs(P, Q)$, $gr(P, A)$, $g(P, Q)$, $g(P, A)$, $g(A, B)$, where P and Q are points of Ψ and A and B are points of U, are used in connection with oriented geodesic segments, geodesic rays, or geodesics, the orientation is such that the order of the two determining points is that given.

THEOREM 9.1. *Any element of a right (left) horocycle $C(P, A)$ is the limit element of elements on right (left) circles through points approaching P and with centers approaching A. Conversely, if an element is a limit element of elements on such right (left) circles, it is an element of the right (left) horocycle $C(P, A)$.*

Let q be the element of $C_R(P, A)$ at the point Q. The direction of q is then perpendicular to the direction of the initial element e of $gr(P, A)$ and the relation of e to q is the same as that of the positive x-axis to the positive y-axis. Let C_n $(n = 1, 2, \cdots)$ be a sequence of circles with centers R_n $(n = 1, 2, \cdots)$ such that $\lim_{n\to\infty} R_n = A$ and let $P_n \subset C_n$ $(n = 1, 2, \cdots)$ such that $\lim_{n\to\infty} P_n = P$. It follows from Theorem 7.2 that there exists a sequence of points Q_n, $Q_n \subset C_n$, such that $\lim_{n\to\infty} Q_n = Q$. The initial element e_n of $gs(Q_n, R_n)$ approaches e as n becomes infinite. If C_n is oriented so that it is a right circle, the element q_n of it at Q_n is perpendicular to e_n and e_n has the same relation to q_n that the positive x-axis has to the positive y-axis. It follows that $\lim_{n\to\infty} q_n = q$, and the first part of the theorem is proved in the case of right horocycles and right circles. The proof in the case of left is similar.

Let q at the point Q be an element satisfying the condition with respect to right circles stated in the second part of the theorem. It immediately follows from Theorem 7.2 that $Q \subset C(P, A)$. If the right circles are C_n, with centers R_n, and $Q_n \subset C_n$ such that $\lim_{n\to\infty} Q_n = Q$, it is easily seen that the element q_n of C_n at Q_n must approach q. Similar results apply to the case of left circles, and the proof of the theorem is complete.

The point P divides the horocycle $C(P, A)$ into two parts, one part consisting of P and those points of $C(P, A)$ on one side of $g(P, A)$, the other part consisting of P and the points of $C(P, A)$ on the other side of $g(P, A)$. Each of these parts will be termed a *semihorocycle*. In the notation of §6, one of these semihorocycles is given by $y = y^+(x)$, the other by $y = y^-(x)$. The point A will be called the *point at infinity* of either semihorocycle.

If $C(P, A)$ is oriented, the two semihorocycles have an orientation induced in them, and one of these will have its initial point at P. If $C(P, A)$ is so oriented that it becomes a right (left) horocycle, that semihorocycle which has its initial point at P will be called a *right* (*left*) *semihorocycle* and will be denoted by $SC_R(P, A)$ $\{SC_L(P, A)\}$. Again if the notation of §6 is used, $SC_R(P, A)$ is given by $y = y^+(x)$, with initial point P, and $SC_L(P, A)$ is given by $y = y^-(x)$, with initial point P.

THEOREM 9.2. *The right* (*left*) *semihorocycles with initial point P form a field in Ψ except at P.*

Let Q be any point of Ψ other than P. Then according to Theorem 8.1 there are just two horocycles containing both P and Q and the points at infinity A and B of these horocycles are the points at infinity of the set $E(P, Q)$. The point P divides each of these horocycles into two semihorocycles of which just one contains Q. Thus there are just two oriented semihorocycles with initial point P which contain Q. To complete the proof it is sufficient to show that one of these is a right and the other a left semihorocycle.

Neither A nor B can coincide with the points at infinity of $g(P, Q)$. We can assume that the notation has been so chosen that if the initial element of $gr(P, A)$ is rotated in the clockwise direction until it coincides with the initial element of $gr(P, Q)$, the angle of rotation is less than π. It was shown in §5 that A and B lie on opposite sides of $g(P, Q)$. Therefore if the initial element of $gr(P, B)$ is rotated in the counterclockwise direction until it coincides with the initial element of $gr(P, Q)$, the angle of rotation is less than π. But then in the normal geodesic coördinate system with P as origin and $gr(P, A)$ as positive x-axis, Q has a negative y-coördinate, while in the normal geodesic coördinate system with P as origin and $gr(P, B)$ as positive x-axis, Q has a positive y-coördinate. Thus one of the oriented semihorocycles which have initial point P and which contain Q is a right and the other a left semihorocycle. The proof is complete.

Let R be the point in which a horocycle with point at infinity A intersects the geodesic $g(O, A)$ determined by the origin O and the point A. Let distance on $g(O, A)$ be measured from O and negative in the direction of A. Then R has a coördinate r. Conversely, given A and r, the horocycle is completely determined, for it is the horocycle $C(R, A)$. The number r will be called the *radius* of the horocycle and the horocycle determined by r and A will be denoted by $C(r, A)$.

We may state without proof that the circle with center O and radius $|r|$ contains no points of the horocycle $C(r, A)$ in its interior.

THEOREM 9.3. *Let $C_R(r_n, A_n)$ $(n = 1, 2, \cdots)$ be a sequence of right horocycles such that $\lim_{n\to\infty} r_n = r$ and $\lim_{n\to\infty} A_n = A$. If p is an arbitrary element of $C_R(r, A)$, there exist elements p_n of $C_R(r_n, A_n)$ $(n = 1, 2, \cdots)$ such that $\lim_{n\to\infty} p_n = p$. The corresponding result obtained when right is replaced by left holds.*

Let R be the point where $C_R(r, A)$ cuts $g(O, A)$ and let R_n be the point where $C_R(r_n, A_n)$ cuts $g(O, A_n)$. Since, with increasing n, r_n approaches r and any finite segment of $g(O, A)$ is approximated uniformly closely by $g(O, A_n)$, necessarily $\lim_{n\to\infty} R_n = R$. But then, from Theorem 7.3, if P is the point bearing p, there exist points P_n $(n = 1, 2, \cdots)$, $P_n \subset C_R(r_n, A_n)$, such that $\lim_{n\to\infty} P_n = P$. It follows that if e_n denotes the initial element of $gr(P_n, A_n)$, $\lim_{n\to\infty} e_n = e$, the initial element of $gr(P, A)$. The element p_n of $C_R(r_n, A_n)$ at P_n is obtained from e_n by rotating e_n through the angle $\frac{1}{2}\pi$. Since p is obtained from e by a rotation of $\frac{1}{2}\pi$, it follows that $\lim_{n\to\infty} p_n = p$, and this is the stated result. The similar proof for left horocycles is omitted.

THEOREM 9.4. *If the sequence of elements p_n $(n = 1, 2, \cdots)$ approaches the element p as n becomes infinite, and if the right (left) horocycles determined by p_n and p have points at infinity A_n and A and radii r_n and r, respectively, $\lim_{n\to\infty} A_n = A$ and $\lim_{n\to\infty} r_n = r$.*

The proof will be carried through only for the case of right horocycles. The proof for left horocycles is entirely similar.

Let P_n be the point bearing p_n and let P be the point bearing p. Then the initial element e_n of the directed geodesic ray $gr(P_n, A_n)$ is perpendicular to p_n and the relation of e_n to p_n is the same as that of the positive x-axis to the positive y-axis. The same relationship holds between e, the initial element of $g(P, A)$ and p. Since $\lim_{n\to\infty} p_n = p$, it follows that $\lim_{n\to\infty} e_n = e$ and consequently $\lim_{n\to\infty} A_n = A$.

The set $| r_n |$ is a bounded set. For the circle with center at the origin O and radius r_n does not contain P_n in its interior and consequently $| r_n | \leqq OP_n$. Since $OP_n \leqq OP + PP_n$, and the set PP_n is bounded, it follows that the set $| r_n |$ $(n = 1, 2, \cdots)$ is a bounded set.

Let Q_n be the point in which $C(P_n, A_n)$ intersects $g(O, A_n)$ and let Q be the point where $C(P, A)$ intersects $g(O, A)$. Then $\lim_{n\to\infty} Q_n = Q$. For since the set $| r_n | = OQ_n$ is bounded and $\lim_{n\to\infty} A_n = A$, the only limit points of the set Q_n must be on $g(O, A)$. Suppose that there exists a subsequence Q_{n_i} $(i = 1, 2, \cdots)$ such that $\lim_{i\to\infty} Q_{n_i} = Q^* \neq Q$. Then Q^* is on $g(O, A)$, and since Q is the only point of $C(P, A)$ on $g(O, A)$, Q^* is not on $C(P, A)$. But then, as in the proof of Theorem 7.2, A is not a point at infinity of $E(P, Q^*)$ and there exists a neighborhood η of A such that for any point S of Ψ in η $| SP - SQ^* |$ exceeds a positive constant, independent of S. Since $\lim_{n\to\infty} P_n = P$, $\lim_{n\to\infty} Q_{n_i} = Q^*$ and $\lim_{n\to\infty} A_n = A$, there exist points S in η for which $| SP - SQ^* |$

is less than a given constant. Therefore Q^* must coincide with Q, and so $\lim_{n\to\infty} Q_n = Q$. It follows at once that $\lim_{n\to\infty} r_n = r$.

THEOREM 9.5. *Let $C(r_n, A_n)$ $(n = 1, 2, \cdots)$ be a sequence of horocycles such that $\lim_{n\to\infty} A_n = A$ and $\lim_{n\to\infty} r_n = +\infty$ and let B and D be points of U distinct from A. Then there exists an $\bar{n}$ such that for $n > \bar{n}$, $C(r_n, A_n)$ intersects $g(B, D)$ in two points. As n becomes infinite, one of the points approaches B, the other D, and the angle of intersection at each of these points approaches $\frac{1}{2}\pi$.*

Let Δ be an interval of U with A as midpoint and such that neither B nor D is in Δ. Let $\bar{n}$ be so chosen that for $n > \bar{n}$, A_n is in Δ and r_n is greater than the minimum distance from O to $g(B, D)$. Then for $n > \bar{n}$, the circle with center O and radius r_n intersects $g(B, D)$ in two points P_n and Q_n. Since all the interior points of the segment P_nQ_n are interior points of this circle, they are interior points of $C(r_n, A_n)$. Since B and D are exterior points of $C(r_n, A_n)$, assuming that the order of points on $g(B, D)$ is BP_nQ_nD, we see that each of the geodesic rays P_nB and Q_nD intersects $C(r_n, A_n)$. Let these points of intersection be S_n and T_n, respectively. Since $\lim_{n\to\infty} r_n = +\infty$, the segment P_nQ_n becomes infinite as n becomes infinite and eventually includes any finite segment of $g(B, D)$. Thus $\lim_{n\to\infty} P_n = B$, $\lim_{n\to\infty} Q_n = D$, and consequently $\lim_{n\to\infty} S_n = B$ and $\lim_{n\to\infty} T_n = D$.

As n becomes infinite, the angle at S_n between $gr(S_n, A)$ and $gr(S_n, D)$ approaches zero. Given $\epsilon > 0$, there exists an integer n_1 such that this angle at S_{n_1} is less than $\frac{1}{2}\epsilon$. Let E be so chosen on U—but not on the arc ADB—and so near A that the angle at S_{n_1} between $gr(S_{n_1}, A)$ and $gr(S_{n_1}, E)$ is less than $\frac{1}{2}\epsilon$. For n sufficiently large, the point S_n will lie on $gr(S_{n_1}, B)$ and A_n will lie on the arc DAE. But then $gr(S_n, A_n)$ will cross $gr(S_{n_1}, E)$ in some point V_n. In the geodesic triangle $V_nS_{n_1}S_n$, the angle φ at S_n is less than the exterior angle $ES_{n_1}D$ and this is less than ϵ. The angle φ is the angle between the geodesic rays $gr(S_n, A_n)$ and $gr(S_n, D)$ and thus is less than ϵ. Since $C(r_n, A_n)$ is normal to $gr(S_n, A_n)$ at S_n, the stated result with respect to the angle of intersection at S_n follows. A similar proof holds for the angle of intersection at T_n, and the proof is complete.

10. **Permanent regional transitivity.** Let F be a Fuchsian group with principal circle U and let $\lambda^2(u, v)$ of (2.1) be invariant under F. Then the metric (2.1) will be invariant under F, the corresponding geodesics will be invariant and, since Euclidean angle is invariant under such transformations and angle defined by (2.1) is Euclidean, angle will be invariant.

If points congruent under F are considered identical, there is defined a two-dimensional manifold M of negative curvature which may or may not be closed, depending on the nature of the group F. Just as in the case of constant negative curvature (cf. Hedlund [6], p. 539) the directed geodesics define a flow

in the space E of elements. It was shown in the paper just referred to that in the case of constant negative curvature and Fuchsian group of the first kind the flow has the property of permanent regional transitivity; that is, given any two open sets in the space of elements, either one of these eventually flows into the other. Furthermore, this intersection is permanent in the sense that it holds permanently after a certain time. A further result needed to extend this result to the case of variable negative curvature is the following theorem, which is obvious in the case of constant curvature.

If a geodesic g is periodic on M, there must be a transformation T of F taking g into itself. The points at infinity of g must then be fixed points of T and T is a hyperbolic transformation which leaves g invariant.

THEOREM 10.1. *On a periodic geodesic $g(B, D)$ let there be given a sequence of congruent points P_n ($n = 1, 2, \cdots$) such that $\lim_{n\to\infty} P_n = D$, and let P be any point of Ψ. Then the angle at P_n between $g(B, D)$ and the right (left) semihorocycles with initial point P and passing through P_n approaches $\frac{1}{2}\pi$ as n becomes infinite.*

Let A_n be the point at infinity of $SC_R(P, P_n)$. Then $gr(P_n, A_n)$ is perpendicular to $SC_R(P, P_n)$ at P_n. To prove the theorem it is sufficient to show that if α_n is the angle between the initial elements of $gr(P_n, A_n)$ and $gr(P_n, B)$, $\lim_{n\to\infty} \alpha_n = 0$. As n becomes infinite the angle between the initial elements of $gs(P_n, P)$ and $gr(P_n, B)$ approaches zero, so that $\lim_{n\to\infty} \alpha_n = 0$ if $\lim_{n\to\infty} \beta_n = 0$, where β_n is the angle between the initial elements of $gs(P_n, P)$ and $gr(P_n, A_n)$.

Since $g(B, D)$ is periodic and the points P_n are congruent, there exists a transformation T of F such that under $T^{N(n)}$ the points P_n can be carried back to a fixed point S of $g(B, D)$. Since $\lim_{n\to\infty} P_n = D$, either $\lim_{n\to\infty} N(n) = +\infty$ or $\lim_{n\to\infty} N(n) = -\infty$, and the notation is assumed such that the first is the case. Then $\lim_{n\to\infty} T^{N(n)}(P) = B$. We show that if $A'_n = T^{N(n)}(A_n)$, $\lim_{n\to\infty} A'_n = B$.

If this were not the case, there would exist a sequence n_i ($i = 1, 2, \cdots$) of integers such that $\lim_{i\to\infty} A'_{n_i} = C$, where C is not identical with B. Let $\overset{\frown}{GH}$ be a closed interval of U with C as center and so small that B is an exterior point of it. Then $SC_L(S, G)$, $SC_L(S, H)$, and $\overset{\frown}{GH}$ form a simple closed curve bounding a region of which B is an exterior point. There exists a neighborhood of B such that any left semihorocycle with initial point S and point at infinity in $\overset{\frown}{GH}$ has no point in this neighborhood. Since there exists an I such that for $i > I$, A'_{n_i} is in $\overset{\frown}{GH}$, all the left horocycles $T^{N(n_i)}SC_L(P_n, A_n)$, $i > I$, are such that there is no point of any one of them in the stated neighborhood of B. But $T^{N(n)}(P)$ is on $T^{N(n)}SC_L(P_n, A_n)$ and the fact that $\lim_{n\to\infty} T^{N(n)}(P) = B$ is contradicted. We conclude that $\lim_{n\to\infty} A'_n = B$.

Since T and its powers preserve angle, the angle β_n is equal to the angle between the initial elements of $gs(S, T^{N(n)}(P))$ and $gr(S, A'_n)$. Since $\lim_{n\to\infty} A'_n = B$ and $\lim_{n\to\infty} T^{N(n)}(P) = B$, we conclude that the last angle approaches zero and thus $\lim_{n\to\infty} \beta_n = 0$. This completes the proof of the theorem.

The theorems which allow the application of the methods of Hedlund [6] in the case of constant negative curvature to the case of variable negative curvature are now at hand. The following theorems of Hedlund hold for a manifold of variable negative curvature and a Fuchsian group of the first kind.

THEOREM 1.1′. *If the group F is a Fuchsian group of the first kind with principal circle U, P is an arbitrary point of Ψ, and AB is an arbitrary interval of U, then there exist points C and D of AB such that $SC_L(P, C)$ and $SC_R(P, D)$ are both transitive.*

THEOREM 1.2′. *If F is a Fuchsian group of the first kind, there exists an infinite set of transitive directed horocycles through any point of Ψ. The points at infinity of these transitive horocycles form an everywhere dense set on U.*

THEOREM 2.1′. *If one directed horocycle with A as point at infinity is transitive, all the directed horocycles with A as point at infinity are transitive.*

THEOREM 2.2′. *If F is a Fuchsian group of the first kind, the end points of all axes of hyperbolic transformations of F are h-transitive.*

THEOREM 2.3′. *If F is a Fuchsian group of the first kind and there are copies of the horocycle $C(r, A)$ with radii arbitrarily large, A is h-transitive.*

THEOREM 2.4′. *Let F be a Fuchsian group of the first kind, A a point of U and $gr(O, A)$ the geodesic ray with origin O as initial point and with A as point at infinity. If there exists on $gr(O, A)$ a sequence of points $O_0, O_1, O_2, \cdots$ such that $\lim_{n\to\infty} OO_n = +\infty$ and such that O_n has a copy O'_n $(n = 0, 1, 2, \cdots)$ with OO'_n bounded, n arbitrary, A is h-transitive.*

THEOREM 2.5′. *If F has a fundamental region R which together with its boundary lies entirely interior to U, all points of U are h-transitive.*

THEOREM 2.6′. *If F is of the first kind and if the only boundary points of R on U are parabolic points, all points of U, with the exception of those which are fixed points of parabolic transformations of F, are h-transitive.*

THEOREM 3.1′. *If F is of the first kind, the flow defined by the geodesics has the property of permanent regional transitivity.*

BIBLIOGRAPHY

1. L. BIEBERBACH, *Über Tchebychefsche Netze auf Flächen negativer Krümmung, sowie auf einigen weiteren Flächenarten*, Sitzungsberichte der Preussischen Akademie der Wissenschaften, 1926, XXIII, pp. 294–300.
2. G. A. BLISS, *The Calculus of Variations*, Carus Monograph, Chicago, 1925.

3. H. Busemann, *Über die Geometrien, in denen die Kreise mit unendlichem Radius die kürzesten Linien sind*, Mathematische Annalen, vol. 106(1932), pp. 140–160.

4. J. Hadamard, *Les surfaces à courbures opposées et leur lignes géodésiques*, Journal de Mathématiques pures et appliquées, (5), vol. 4(1898), pp. 27–73.

5. G. A. Hedlund, *Two-dimensional manifolds and transitivity*, Annals of Mathematics, vol. 37(1936), pp. 534–542.

6. G. A. Hedlund, *Fuchsian groups and transitive horocycles*, this Journal, vol. 2(1936), pp. 530–542.

7. G. A. Hedlund, *The dynamics of geodesic flows*, Bulletin of the American Mathematical Society, vol. 45 (1939), pp. 241–260.

8. H. Hopf and W. Rinow, *Ueber den Begriff der vollständigen differentialgeometrischen Fläche*, Commentarii Mathematici Helvetici, vol. 3(1931), pp. 209–225.

9. M. Morse, *A fundamental class of geodesics on any closed surface of genus greater than one*, Transactions of the American Mathematical Society, vol. 26(1924), pp. 25–61.

10. M. Morse, *Instability and transitivity*, Journal de Mathématiques pures et appliquées, (9), vol. 14(1935), pp. 49–71.

Bryn Mawr College.

THE MEASURE OF GEODESIC TYPES ON SURFACES OF NEGATIVE CURVATURE

By Gustav A. Hedlund

1. **Introduction.** Various problems arise in connection with the transitivity of a dynamical system. The first of these concerns the existence of transitive motions. Secondly, if there exist such motions, what is the measure of the totality of transitive motions? Thirdly, is the system metrically transitive? If the last holds, almost all the motions are transitive, so that each of the first two of the above properties is a consequence of the following one.

The first of these problems has been solved in the case of the geodesic problem on a class of surfaces of negative curvature and even for surfaces with some positive or zero curvature, provided there is not enough to destroy the instability of the geodesics (cf. Morse [1], Hedlund [1]). The second and third problems have been solved only for a restricted subclass of these surfaces, namely, those of constant negative curvature, of finite area, and of finite connectivity (cf. E. Hopf [1]). The constancy of the curvature plays an important rôle in the proofs of these results, for it implies that certain transformations are analytic and thus transform sets of measure zero into sets of measure zero. It is the lack of information concerning the corresponding transformations in the variable case which causes the difficulty in applying the methods of the constant case.

This paper gives a solution of the second problem for a class of surfaces of negative curvature, not necessarily constant, of finite area and finite connectivity. This class includes, in particular, all closed orientable surfaces of negative curvature. It also includes surfaces with "parabolic" openings. It is shown that on all surfaces of the stated class almost all the geodesics are transitive. The extension of this result to non-orientable surfaces is easily proved.

In the definition of the preceding class of surfaces, a Fuchsian group of the first kind is used. If the defining Fuchsian group is of the second kind, the geodesics behave in an entirely different manner. It is shown that in this case almost all the geodesics are unstable in the sense that for both future and past time they eventually remain outside any fixed finite region of the surface. (Cf. E. Hopf [1] for results of this kind in the case of constant negative curvature.) This class includes surfaces with "hyperbolic" openings similar to the surfaces which Hadamard (cf. Hadamard [1]) constructed. The class does not include all these Hadamard surfaces, however, so that it is not possible to say that in all cases the perfect sets of geodesics discovered by Hadamard are sets of measure zero. This problem will be taken up in a later paper.

The method used is similar to that used by Tuller (cf. Tuller [1]) in attaining

Received August 24, 1938.

similar results for three-dimensional manifolds of constant negative curvature. It was first necessary to extend various results concerning horocycles (cf. Hedlund [2]) to the case of variable negative curvature, work which has been carried through by Grant. (Cf. the preceding paper of this Journal. This paper will hereafter be referred to as Grant.) The present paper derives the added properties which are needed to solve the stated problems.

2. **A class of simply-connected two-dimensional manifolds of negative curvature.** Let U be the unit circle, $u^2 + v^2 = 1$, and let Ψ be its interior. The metric in Ψ will be defined by

$$ds^2 = \frac{\lambda^2(u, v)(du^2 + dv^2)}{(1 - u^2 - v^2)^2}, \tag{2.1}$$

where $\lambda(u, v)$ is of class C^k, $k \geqq 7$, and

$$0 < a \leqq \lambda(u, v) \leqq b \tag{H.1}$$

in Ψ. The total or Gaussian curvature $K(u, v)$ of (2.1) is then of at least class C^5 in Ψ, and it will be assumed that $\lambda(u, v)$ is such that in Ψ

$$-d^2 \leqq K(u, v) \leqq -c^2 < 0, \tag{H.2}$$

where c and d are positive constants. The condition (H.2) is satisfied, in particular, if $\lambda(u, v)$ is constant, for then the curvature is constant and negative.

As a consequence of the hypothesis of negative curvature, the geodesics have many properties in common with the hyperbolic lines (i.e., the geodesics when λ = const.). There is a unique geodesic segment joining two points P_1 and P_2 of Ψ and this geodesic yields a minimum of arc length with respect to all rectifiable curves lying in Ψ and joining P_1 and P_2. Thus no two geodesics intersect more than once, all geodesic segments are of class A, all unending geodesics are of class A and each is of the type of a hyperbolic line. There is just one unending geodesic of the type of a given hyperbolic line, so the geodesics and hyperbolic lines are in one-to-one correspondence. The geodesics through a point P of Ψ form a field in Ψ except at P. Thus it is possible to set up geodesic polar coördinates with any point P of Ψ as center.

It has been shown by Grant (cf. Grant, §6, et seq.) that, under the conditions imposed, there exist curves called horocycles which display the properties of the horocycles in the case of constant curvature. In particular, a horocycle is a limiting curve of geodesic circles with radii becoming infinite and is an orthogonal trajectory of the geodesics which have a common point at infinity.

The *distance* $D(P_1, P_2)$ between two points P_1 and P_2 of Ψ will be defined as the arc length (as determined by (2.1)) of the unique geodesic segment joining P_1 and P_2.

An *element* $e(u, v, \theta)$, $u^2 + v^2 < 1$, $0 \leqq \theta < 2\pi$, determines a point $P(u, v)$ of Ψ and a direction θ at that point, where the angle θ is to be measured in the counterclockwise sense from a direction parallel to the positive u-axis. Con-

versely, a point of Ψ and a direction at this point determines a unique element. The distance $D_e(e_1, e_2)$ between elements $e_1(u_1, v_1, \theta_1)$ and $e_2(u_2, v_2, \theta_2)$ will be defined by

$$D_e(e_1, e_2) = D(P_1, P_2) + \{\theta_2 - \theta_1\},$$

where P_1 is the point (u_1, v_1), P_2 is the point (u_2, v_2) and $\{\theta_2 - \theta_1\}$ is the least value of the set $|\theta_2 - \theta_1 + 2n\pi|$ $(n = 0, \pm 1, \cdots)$. The space of elements $e(u, v, \theta)$, $u^2 + v^2 < 1$, $0 \leqq \theta < 2\pi$, will be denoted by E. The volume in E will be defined by the integral

$$\int\int\int \frac{\lambda^2(u, v)}{(1 - u^2 - v^2)^2} \, du \, dv \, d\theta, \tag{2.2}$$

and measurability and measure will be that determined by this definition of volume.

If we interpret time as arc length along the geodesics, the geodesics define a flow of the space E into itself. It is easily shown that the integral (2.2) and hence measure are invariant under this flow.

3. **A differential geometric identity.** Since the geodesics through a point P of Ψ form a field in Ψ except at P, it is possible to set up geodesic polar coordinates with P as center. The first fundamental form is then given by

$$ds^2 = dr^2 + G^2(r, \phi) d\phi^2, \tag{3.1}$$

where, because of the hypotheses on the original quadratic form, the non-negative function $G(r, \phi)$ is of at least class C^4 in the region

$$0 \leqq r < +\infty, \qquad -\infty < \phi < +\infty \tag{R}$$

and

$$G(r, \phi + 2n\pi) \equiv G(r, \phi) \qquad (n = 0, \pm 1, \pm 2, \cdots).$$

The following boundary conditions are satisfied

$$G(0, \phi) \equiv 0, \qquad \left(\frac{\partial G}{\partial r}\right)_{r=0} = 1, \tag{3.2}$$

and $G(r, \phi)$ satisfies the differential equation

$$\frac{\partial^2 G}{\partial r^2} + K \cdot G = 0, \tag{3.3}$$

where $K = K(r, \phi)$ is the Gaussian curvature and is of at least class C^4 in R. With the aid of (3.2) and (3.3) we see that

$$G_\phi(0, \phi) = G_{rr}(0, \phi) = G_{r\phi}(0, \phi) = G_{\phi\phi}(0, \phi)$$
$$= G_{rr\phi}(0, \phi) = G_{r\phi\phi}(0, \phi) = G_{\phi\phi\phi}(0, \phi) = 0.$$

Thus

$$G(r, \phi) = r + r^3R(r, \phi), \qquad G_\phi(r, \phi) = r^3R^*(r, \phi) \tag{3.4}$$

in R, where $|R(r, \phi)|$ and $|R^*(r, \phi)|$ are bounded functions in any region $0 \leqq r \leqq \delta$, $-\infty < \phi < +\infty$, where $\delta > 0$.

In the case of constant negative curvature, $G(r, \phi)$ is a function of r alone and the expression for it in terms of exponential functions is readily determined from (3.3). In this case the fundamental identity

$$G_\phi(r, \phi) = -G(r, \phi) \int_0^r \frac{1}{G^2(s, \phi)} \left\{\int_0^s K_\phi(t, \phi)G^2(t, \phi)\, dt\right\} ds, \qquad r > 0, \tag{I}$$

where one of the integrals on the right is an improper integral, is trivial, for both sides vanish identically. We show that the identity holds in the general case under consideration.

To that end, let

$$h(s) = \frac{1}{G^2(s, \phi)} \int_0^s K_\phi(t, \phi)G^2(t, \phi)\, dt, \qquad s > 0.$$

Then $h(s)$ is continuous, $s > 0$. It follows from (3.4) and the boundedness of $R(r, \phi)$ in a neighborhood of $r = 0$ that there exists a $\delta > 0$ such that

$$G^2(r, \phi) > \tfrac{1}{2}r^2, \qquad G^2(r, \phi) < \tfrac{3}{2}r^2, \qquad 0 \leqq r \leqq \delta.$$

Let B be an upper bound of $|K_\phi(r, \phi)|$ for $0 \leqq r \leqq \delta$. Then for $0 < s \leqq \delta$ we have

$$|h(s)| \leqq \frac{2}{s^2} B \int_0^s \frac{3t^2}{2}\, dt = Bs.$$

It follows that $\lim_{s \to 0} h(s) = 0$ and consequently the integral on the right in (I) exists.

If we differentiate (3.3) partially with respect to ϕ, there results

$$G_{rr\phi} = -K_\phi G - KG_\phi,$$

and thus

$$\int_0^s K_\phi(t, \phi)G^2(t, \phi)\, dt = -\int_0^s \{G(t, \phi)G_{tt\phi}(t, \phi) - G_{tt}(t, \phi)G_\phi(t, \phi)\}\, dt.$$

By integration by parts, we obtain

$$\int_0^s K_\phi(t, \phi)G^2(t, \phi)\, dt = -G(s, \phi)G_{s\phi}(s, \phi) + G_s(s, \phi)G_\phi(s, \phi),$$

and thus

$$h(s) = \frac{1}{G^2(s, \phi)} \{-G(s, \phi)G_{s\phi}(s, \phi) + G_s(s, \phi)G_\phi(s, \phi)\} = -\frac{\partial}{\partial s}\left[\frac{G_\phi(s, \phi)}{G(s, \phi)}\right].$$

Thus the right side of (I) is equal to

$$G_\phi(r, \phi) - G(r, \phi) \lim_{\epsilon \to 0} \frac{G_\phi(\epsilon, \phi)}{G(\epsilon, \phi)}.$$

Making use of (3.4), we have

$$\frac{G_\phi(\epsilon, \phi)}{G(\epsilon, \phi)} = \frac{\epsilon^2 R^*(\epsilon, \phi)}{1 + \epsilon^2 R(\epsilon, \phi)},$$

and since $|R^*(\epsilon, \phi)|$ and $|R(\epsilon, \phi)|$ are bounded in $0 \leqq \epsilon \leqq \delta$, $\delta > 0$, ϕ arbitrary,

$$\lim_{\epsilon \to 0} \frac{G_\phi(\epsilon, \phi)}{G(\epsilon, \phi)} = 0.$$

The proof that the identity (I) holds is complete.

4. **The oscillation of $G(r, \phi)$ on arcs of geodesic circles.** If G_M and G_m denote the maximum and minimum, respectively, of $G(r, \phi)$ on an arc of length l of a geodesic circle, the ratio G_M/G_m will in general vary with l, the arc, and the geodesic circle. In the case of constant curvature, G depends only on r, so that the ratio always has the value 1. It will be shown that if an additional hypothesis, which is automatically fulfilled in a large number of cases, is made, the ratio G_M/G_m is less than a constant which depends only on l and not on the particular geodesic circle or the arc of it chosen.

The additional hypothesis concerns the curvature and is as follows.

(H.3) *There exists a constant A such that*

$$\left| \frac{1}{G} \cdot \frac{\partial K}{\partial \phi} \right| < A, \qquad r > 0,$$

where the geodesic polar coördinate system under consideration is arbitrary.

If s is the arc length on the geodesic circle passing through the point (r, ϕ), $r > 0$, it follows from (3.1) that $d\phi/ds = G^{-1}$, and thus

$$\frac{1}{G} \cdot \frac{\partial K}{\partial \phi} = \frac{dK}{ds},$$

where dK/ds is the directional derivative of K along the geodesic circle. Thus if $|dK/ds|$ is uniformly bounded by a constant which is independent of the geodesic polar coördinate system used, the condition (H.3) is satisfied.

The first step in the derivation of the desired result is the proof that there exists a constant B, independent of the polar coördinate system chosen, such that

$$\left| \frac{1}{G^2} \cdot \frac{\partial G}{\partial \phi} \right| \leqq B, \qquad r \geqq 1.$$

It follows from the identity (I) of §3 that

$$\frac{1}{G^2}\cdot\frac{\partial G}{\partial \phi} = -\frac{1}{G(r,\phi)}\int_0^r \frac{1}{G^2(s,\phi)}\left\{\int_0^s K_\phi(t,\phi)G^2(t,\phi)\,dt\right\}ds,$$

and hence, with the aid of (H.3), we obtain

$$\left|\frac{1}{G^2}\cdot\frac{\partial G}{\partial \phi}\right| \leqq \frac{A}{G(r,\phi)}\int_0^r \frac{1}{G^2(s,\phi)}\left\{\int_0^s G^3(t,\phi)\,dt\right\}ds. \tag{4.1}$$

In order to study the function on the right of inequality (4.1) we make considerable use of the fact that $G(r, \phi)$ satisfies (3.3). It is somewhat simpler to consider the ordinary differential equation obtained from (3.3) by holding ϕ constant.

Let $y(x)$ be that solution of the differential equation

$$y'' = f(x)\cdot y, \tag{4.2}$$

$f(x)$ continuous and $0 < c^2 \leqq f(x) \leqq d^2$ in $0 \leqq x < +\infty$, determined by the initial conditions

$$y(0) = 0, \qquad y'(0) = 1. \tag{4.3}$$

Then $y(x)$ is of class C″, $0 \leqq x < +\infty$, and since $f(x)$ is positive, it follows that $y(x) > 0$ and $y'(x) > 0$, $0 < x < +\infty$. With the aid of well known comparison methods, we have

$$c\,\frac{e^{cx}+e^{-cx}}{e^{cx}-e^{-cx}} \leqq \frac{y'(x)}{y(x)} \leqq d\,\frac{e^{dx}+e^{-dx}}{e^{dx}-e^{-dx}}, \qquad x > 0. \tag{4.4}$$

Since the function on the left is decreasing for $x > 0$ and approaches c as limit when x becomes positively infinite, it follows that

$$\frac{y'(x)}{y(x)} > c, \qquad x > 0. \tag{4.5}$$

Since the function on the right is decreasing, it follows that

$$\frac{y'(x)}{y(x)} \leqq d\,\frac{e^{d}+e^{-d}}{e^{d}-e^{-d}}, \qquad x \geqq 1. \tag{4.6}$$

Now consider the function

$$v(x) = \frac{1}{y^3}\int_0^x y^3\,dx, \qquad x > 0,$$

where $y(x)$ is again the solution of (4.2) satisfying the initial conditions (4.3). Then $v(x)$ is of class C″ and positive for $x > 0$.

LEMMA 4.1. *There exist positive constants d_1 and d_2, depending only on c and d, such that*

$$d_1 < v(x) < d_2, \qquad x \geqq 1. \tag{4.7}$$

By comparison we have

$$\frac{e^{cx} - e^{-cx}}{2c} \leqq y(x) \leqq \frac{e^{dx} - e^{-dx}}{2d}, \tag{4.8}$$

and thus

$$0 < c_1 = \frac{8d^3}{(e^d - e^{-d})^3} \int_0^1 \left(\frac{e^{cx} - e^{-cx}}{2c}\right)^3 dx \leqq v(1)$$

$$\leqq \frac{8c^3}{(e^c - e^{-c})^3} \int_0^1 \left(\frac{e^{dx} - e^{-dx}}{2d}\right)^3 dx = c_2.$$

Let d_1 be the smaller of the two positive constants $\frac{1}{2}c_1$ and

$$\frac{e^d + e^{-d}}{6d(e^d - e^{-d})}.$$

Let d_2 be the greater of $2c_2$ and $\frac{1}{2}c^{-1}$. The constants d_1 and d_2 depend only on c and d.

By differentiation it follows that $v(x)$ is a solution of the differential equation

$$v' + v\frac{3y'}{y} = 1, \qquad x > 0.$$

At $x = 1$ the inequalities of the lemma are satisfied. Suppose that $x_0 > 1$ and $v(x_0) = d_1$. We can suppose that x_0 is the least of such values and under this assumption $v'(x_0) \leqq 0$, since $v(1) \geqq c_1 > d_1$. But with the aid of (4.6),

$$v'(x_0) = 1 - v(x_0)\frac{3y'(x_0)}{y(x_0)} \geqq 1 - d_1 \cdot 3d\frac{e^d - e^{-d}}{e^d + e^{-d}} > 0,$$

and from this contradiction we infer that $v(x_0) \neq d_1$, $x_0 \geqq 1$. Since $v(x)$ is continuous and $v(1) \geqq c_1 > d_1$, it follows that $v(x) > d_1$, $x \geqq 1$.

Suppose that $x_0 > 1$ and $v(x_0) = d_2$. Assuming that x_0 is the least of such values, since $v(1) = c_2 < d_2$, we have $v'(x_0) \geqq 0$. But from (4.5)

$$v'(x_0) = 1 - v(x_0)\frac{3y'(x_0)}{y(x_0)} \leqq 1 - d_2 3c < 0.$$

Thus $v(x_0) \neq d_2$, $x_0 > 1$. Since $v(x)$ is continuous and $v(1) \leqq c_2 < d_2$, it follows that $v(x) < d_2$, $x \geqq 1$, and the proof of the lemma is complete.

Let the function $z(x)$ be defined as follows:

$$\begin{cases} z(x) = \dfrac{1}{y^2(x)} \displaystyle\int_0^x y^3(\eta)\, d\eta, & x > 0, \\ z(0) = 0, \end{cases}$$

where $y(x)$ is again the solution of (4.2) satisfying the initial conditions (4.3). The function $z(x)$ is continuous for $x \geqq 0$. The continuity for $x > 0$ follows

from the continuity of $y(x)$, $x > 0$, and the fact that $y(x) > 0$, $x > 0$. To show the continuity at $x = 0$, we make use of (4.8) and obtain

$$0 < z(x) \leqq \frac{4c^2}{(e^{cx} - e^{-cx})^2} \int_0^x \frac{(e^{d\eta} - e^{-d\eta})^3}{8d^3} \, d\eta, \qquad x > 0,$$

whence

$$0 < z(x) \leqq \frac{4c^2}{(e^{cx} - e^{-cx})^2} \frac{(e^{dx} - e^{-dx})^3}{8d^3} x, \qquad x > 0.$$

Since the function on the right approaches zero as x approaches zero, $\lim_{x \to 0} z(x) = 0$, and $z(x)$ is continuous at $x = 0$. It is evident that $z(x)$ is of class C$''$ for $x > 0$.

Let

$$w(x) = \frac{1}{y(x)} \int_0^x z(\xi) \, d\xi, \qquad x > 0,$$

where $y(x)$ and $z(x)$ are as defined above. Then $w(x)$ is of class C$''$ and positive for $x > 0$.

LEMMA 4.2. *There exist positive constants e_1 and e_2, depending only on c and d, such that*

$$e_1 < w(x) < e_2, \qquad x \geqq 1.$$

The method of proof is similar to that used in proving Lemma 4.1, and only the proof of the second inequality will be given. It follows from (4.8) and the definition of $w(x)$ that

$$w(1) \leqq \frac{2c}{e^c - e^{-c}} \int_0^1 \frac{(2c)^2}{(e^{c\xi} - e^{-c\xi})^2} \left\{ \int_0^\xi \left(\frac{e^{d\eta} - e^{-d\eta}}{2d} \right)^3 d\eta \right\} d\xi = h_2 > 0.$$

Let e_2 be the larger of the two positive numbers $2h_2$ and $2d_2/c$, where d_2 is that of Lemma 4.1. Then $w(1) \leqq h_2 < e_2$. Suppose that at $x_0 > 1$, $w(x_0) = e_2$. It can be assumed that x_0 is the least of such values and thus $w'(x_0) \geqq 0$. But it is readily shown that $w(x)$ is a solution of the differential equation

$$w' + w \frac{y'}{y} = \frac{1}{y^3} \int_0^x y^3(\eta) \, d\eta = v(x), \qquad x > 0,$$

and thus, with the aid of (4.5) and Lemma 4.1, we get

$$w'(x_0) = -w(x_0) \frac{y'(x_0)}{y(x_0)} + v(x_0) \leqq -e_2 c + d_2 < 0.$$

From this contradiction we infer that we cannot have $w(x) = e_2$, $x \geqq 1$. But since $w(1) < e_2$ and $w(x)$ is continuous, it follows that $w(x) < e_2$, $x \geqq 1$. The constant d_2 depends only on c and d, so that the same is true of e_2.

LEMMA 4.3. *There exists a constant B, determined by A, c and d, such that*

$$\left| \frac{1}{G^2} \cdot \frac{\partial G}{\partial \phi} \right| \leqq B, \qquad r \geqq 1.$$

For any fixed ϕ, $G(r, \phi)$ is a solution of the differential equation

$$\frac{d^2 y}{dr^2} = -K(r, \phi) \cdot y, \qquad r \geqq 0, \tag{4.9}$$

and satisfies the initial conditions

$$G(0, \phi) = 0, \qquad G_r(0, \phi) = 1. \tag{4.10}$$

Since $0 < c^2 \leqq -K(r, \phi) \leqq d^2$, the equation (4.9) satisfies the conditions on (4.2) and the initial conditions (4.10) are the same as those of (4.3). If we replace $y(x)$ by $G(r, \phi)$, $w(x)$ becomes

$$\frac{1}{G(r, \phi)} \int_0^r \frac{1}{G^2(s, \phi)} \left\{ \int_0^s G^3(t, \phi)\, dt \right\} ds,$$

and it follows from Lemma 4.2 that this function does not exceed e_2 when $r \geqq 1$. If we combine this with (4.1), there results

$$\left| \frac{1}{G^2} \cdot \frac{\partial G}{\partial \phi} \right| \leqq A e_2, \qquad r \geqq 1.$$

If we set $B = Ae_2$, the stated lemma is proved.

We come to the proof of the desired theorem.

THEOREM 4.1. *Given $l > 0$, there exists a constant D, depending only on B and l, such that if any point P in Ψ is used as center of geodesic polar coördinates, with resulting quadratic form*

$$ds^2 = dr^2 + G^2(r, \phi)\, d\phi^2,$$

if σ is an arc of length l of any geodesic circle of radius greater than unity of this system, and $G_M(\sigma)$ and $G_m(\sigma)$ are the maximum and minimum, respectively, of $G(r, \phi)$ on σ, then

$$\frac{G_M}{G_m} < D.$$

Case I. $Bl < 1$.

Let s be the arc length on σ measured from one end point. Then on σ, $G(r, \phi)$ is a function of class C′ of s, and if s_M and s_m denote the values of s where G assumes the values G_M and G_m, respectively, we have

$$G_M - G_m = \left(\frac{dG}{ds} \right)_P (s_M - s_m) = \left(\frac{1}{G} \cdot \frac{\partial G}{\partial \phi} \right)_P (s_M - s_m),$$

where P is a point of σ. With the aid of Lemma 4.3, we obtain

$$G_M - G_m \leqq \left| \frac{1}{G^2} \cdot \frac{\partial G}{\partial \phi} \right|_P \cdot G_P \cdot | s_M - s_m | \leqq BlG_M.$$

It follows that

$$\frac{G_M}{G_m} \leqq \frac{1}{1 - Bl},$$

and if we let $D = (1 - Bl)^{-1}$, the stated result is proved.

Case II. $Bl \geqq 1$.

Let $l_1, l_2, \cdots, l_n$ be a set of positive numbers such that $\sum_{i=1}^{n} l_i = l$ and $Bl_i < 1$ $(i = 1, 2, \cdots, n)$. Let the arc σ be divided into a set of n successive closed arcs $I_1, I_2, \cdots, I_n$, of lengths $l_1, l_2, \cdots, l_n$, respectively. Let ${}_iG_M$ be the maximum of G on I_i and let ${}_iG_m$ be the minimum on this same arc. There exists an integer j, $1 \leqq j \leqq n$, such that $G_M = {}_jG_M$, and an integer k, $1 \leqq k \leqq n$, such that $G_m = {}_kG_m$. We can assume that $j \leqq k$, for the case $j > k$ can be reduced to this case by taking the arcs in the opposite order on σ. Then

$$\frac{G_M}{G_m} = \frac{{}_jG_M}{{}_kG_m} = \frac{{}_jG_M}{{}_jG_m} \cdot \frac{{}_{j+1}G_M}{{}_{j+1}G_m} \cdot \ \cdots \ \cdot \frac{{}_kG_M}{{}_kG_m} \cdot \frac{{}_jG_m}{{}_{j+1}G_M} \cdot \ \cdots \ \cdot \frac{{}_{k-1}G_m}{{}_kG_M},$$

where there is just one factor on the right of $k = j$. Since the intervals I_{j+r} and I_{j+r+1} are closed and adjacent, they have a point in common and

$$\frac{{}_{j+r}G_m}{{}_{j+r+1}G_M} \leqq 1.$$

Applying this and using the fact that Case I applies to each I_i $(i = 1, 2, \cdots, n)$, we see that

$$\frac{G_M}{G_m} \leqq \frac{1}{1 - Bl_j} \cdot \frac{1}{1 - Bl_{j+1}} \cdot \ \cdots \ \cdot \frac{1}{1 - Bl_k}.$$

If we let

$$D = \frac{1}{1 - Bl_1} \cdot \frac{1}{1 - Bl_2} \cdot \ \cdots \ \cdot \frac{1}{1 - Bl_n},$$

the stated result is proved.

5. **The approximation of horocycles by geodesic circles.** Let P be a point of Ψ, A a point of U, and $C(P, A)$ the horocycle determined by P and A. (For the definition, existence and properties of horocycles, cf. Grant.) Let Q be a point of $C(P, A)$. It has been shown by Grant (Theorem 7.2) that if $C_1, C_2, \cdots$ is a sequence of geodesic circles with centers approaching A, and if C_n has on it a point P_n such that $\lim_{n\to\infty} P_n = P$, then C_n has on it a point Q_n such that $\lim_{n\to\infty} Q_n = Q$. A further property of this approximating process will be derived by

showing that the length of the smaller of the arcs P_nQ_n of C_n remains bounded as n becomes infinite. To that end we first prove the following lemma.

LEMMA 5.1. *Given a geodesic circle C, there exists a constant b such that the length of the arc which lies within C of any geodesic circle with center exterior to C does not exceed b.*

If geodesic polar coördinates are set up with the point $P(u, v)$ of Ψ as center, the function $G(r, \phi)$ in the resulting quadratic form

$$dr^2 + G^2(r, \phi)\, d\phi^2$$

will depend on the choice of P. We indicate this by writing $G(r, \phi, u, v)$ instead of $G(r, \phi)$. By application of the theorems concerning the dependence of solutions of differential equations on the initial conditions, it can be shown that $G(r, \phi, u, v)$ depends continuously on the coördinates of P and the choice of initial direction. It follows that if P is restricted to lying on C and $\bar{r}$ is a positive constant, there exists a constant $h(\bar{r})$ such that

$$G(r, \phi, u, v) \leqq h(\bar{r}), \qquad (u, v) \text{ on } C, \qquad 0 \leqq r \leqq \bar{r}. \tag{5.1}$$

Let T be the center and t the radius of C. Let S, any point of Ψ exterior to C, be the center of a geodesic circle C' which cuts across C in points V and W. The segment VW of C' which lies within and on C intersects the geodesic determined by T and S in one point X (for the points V and W lie on opposite sides of the geodesic (cf. Grant, Theorem 4.4) and the geodesic intersects C' in just two points) and we consider the segment XV thus determined. Let M be the point in which the geodesic segment joining T and S intersects C. Then M is interior to C', for the radius of C' must be greater than the length of SM if C' is to intersect C in two points. If we set up geodesic polar coördinates with M as center, XV is given by a function of the form $r = r(\phi)$, $\phi_1 \leqq \phi \leqq \phi_2$, where $r(\phi)$ is an increasing or decreasing function of ϕ (this is easily proved), and we can assume that $0 \leqq \phi_1 < \phi_2 \leqq 2\pi$. Furthermore, since every point of XV lies within or on C, its distance from M cannot exceed $2t$. Let $h(2t)$ be the constant as determined in (5.1). Then the length of XV is given by

$$L = \int_{\phi_1}^{\phi_2} \sqrt{\left(\frac{dr}{d\phi}\right)^2 + G^2(r, \phi)}\, d\phi,$$

where $G(r, \phi)$ is $G(r, \phi, u, v)$ when (u, v) are the coördinates of M. It follows that

$$L \leqq \int_{\phi_1}^{\phi_2} \left[\left| \frac{dr}{d\phi} \right| + G(r, \phi) \right] d\phi,$$

and since $dr/d\phi$ does not change sign, with the aid of (5.1) we obtain

$$L \leqq |\, r(\phi_2) - r(\phi_1) \,| + h(2t)(\phi_2 - \phi_1) \leqq 2t + 2\pi h(2t).$$

A similar result applies to the arc XW of C' and thus the length of the arc VW of C' cannot exceed $4t + 4\pi h(2t)$. If we set b equal to this number, the stated lemma is proved.

The extended approximation theorem is as follows.

THEOREM 5.1. *Let P be a point of Ψ, A a point of U, $C(P, A)$ the horocycle determined by P and A, and Q an arbitrary point of $C(P, A)$. If C_n $(n = 1, 2, \cdots)$ is a sequence of geodesic circles with centers approaching A and C_n has on it a point P_n $(n = 1, 2, \cdots)$ such that $\lim_{n\to\infty} P_n = P$, then there exist a constant L and a point Q_n on C_n $(n = 1, 2, \cdots)$ such that the point Q_n lies on the arc of C_n of length L with midpoint at P_n and $\lim_{n\to\infty} Q_n = Q$.*

The part of the theorem which states that Q_n exists on C_n such that $\lim_{n\to\infty} Q_n = Q$ has been proved by Grant (Grant, Theorem 7.2), so it remains to prove the existence of the constant L with the stated property.

Let C be a geodesic circle with center P and containing Q in its interior. Then there exists an N such that for $n \geqq N$, the center of C_n is exterior to C and P_n and Q_n are interior to C. Let L be so chosen that the arc of length L of C_n $(n = 1, 2, \cdots, N - 1)$ with midpoint P_n contains Q_n and also such that $L > 2b$, where b is the constant determined by C as in Lemma 5.1. Then the stated theorem holds for $n = 1, 2, \cdots, N - 1$. For $n \geqq N$, P_n and Q_n lie within C and the center of C_n is exterior to C. It follows from Lemma 5.1 that the length of the arc P_nQ_n of C_n lying within C cannot exceed b. But since $L > 2b$, the arc of C_n of length L with midpoint at P_n must contain Q_n. The proof of the theorem is complete.

6. **Two-dimensional manifolds of negative curvature and their classification.** Let F be a Fuchsian group with principal circle U. We now impose the following condition on $\lambda(u, v)$.

(H.4) *$\lambda(u, v)$ is invariant under the group F.*

The condition (H.4) implies the invariance of the metric (2.1), and hence the invariance of length, angle, area and curvature as determined by (2.1), as well as the invariance of the geodesics. Two points P_1 and P_2 of Ψ are *congruent* if there is a transformation of the group taking one into the other. Two elements $e_1(u_1, v_1, \theta_1)$ and $e_2(u_2, v_2, \theta_2)$ are *congruent* if $P_1(u_1, v_1)$ and $P_2(u_2, v_2)$ are congruent and there is a transformation of the group F taking P_1 into P_2 such that the direction θ_1 at P_1 is transformed into the direction θ_2 at P_2.

If congruent points are considered identical, there is defined a two-dimensional manifold M of negative curvature. The topological properties of M are determined by the group F. If F contains elliptic transformations, such a transformation having necessarily one fixed point in Ψ and the other fixed point exterior to U, M has one or more singularities of the nature of cusps. That is, the sum of the angles at such a point is less than 2π.

We classify these manifolds according to the properties of F as follows. (For definitions of Fuchsian groups of the first and second kind, cf. Ford, p. 68.)

Class I. *F is of the first kind and has a finite set of generators.*

Class II. *F is of the first kind and has an infinite set of generators.*

Class III. *F is of the second kind.*

Manifolds of class I may or may not be closed. The manifold will be closed if the fundamental region R_0 of F has no boundary points on U. These manifolds include, in particular, all closed, orientable surfaces of negative curvature and such that the coefficients of the first fundamental form are of class C^7. For, from the Gauss-Bonnet formula, such a surface is necessarily of genus greater than one. The universal covering surface of such a surface can be mapped conformally into the interior of U and the resulting quadratic form will be given by (2.1) with λ invariant under a Fuchsian group F, the fundamental region of which has no boundary points on U. Since the surface is closed, it is easily seen that the conditions (H.1), (H.2) and (H.3) are also satisfied.

If, on the other hand, M is of class I but the fundamental region R_0 of F has one or more boundary points on U, M is not closed. In this case F contains a parabolic transformation and M contains a region which is of the topological type of a half-cylinder, of infinite length, and such that a simple closed curve can sweep out this region with its length approaching zero. This is an example of what Hadamard terms a "nappe non évasée" (cf. Hadamard [1], p. 35). We will term such a region a *parabolic opening.*

It is easily shown that the area of a manifold of class I is finite. The area of such a manifold is the area of the fundamental region of the group F defining the manifold.

In the case of manifolds of class III, the fundamental region R_0 contains on its boundary an interval of U (cf. Ford [1], Theorem 14, p. 73). If the end points of such an interval are end points of paired sides of R_0, the manifold M contains a region which is of the topological type of a half-cylinder, of infinite length and such that the length of any closed curve which sweeps out this region becomes infinite. This is an example of a "nappe évasée" of Hadamard (cf. Hadamard [1], p. 35). We will term such a region a *hyperbolic opening.*

A manifold of class III is necessarily of infinite area.

7. **Stability in the sense of Poisson and transitivity.** A motion of a dynamical system is *stable in the sense of Poisson* if the motion returns infinitely often to any given neighborhood of its initial state (cf. Poincaré [1], p. 141). A motion of a dynamical system is *transitive* if the points of the motion are everywhere dense in the phase space. In the case under consideration, namely, that of the geodesics on the manifold M of §6, it will be desirable to formulate these definitions in different, but equivalent terms.

A point e of E determines a point of Ψ and a direction at that point. Thus each element determines a directed geodesic ray having this element as its initial

element. Let gr be this geodesic ray, P its initial point and A its point at infinity. The point e of E, or element, will be said to be ***stable*** (in the sense of Poisson) if there exists a sequence of points $P_1, P_2, \cdots$ on gr, with $\lim_{n\to\infty} P_n = A$, such that if e_n is the element of gr at P_n, there is an element e_n' congruent to e_n such that $\lim_{n\to\infty} e_n' = e$. The element *opposite* e is the element at the same point as e but oppositely directed. The element e will be said to be *completely stable* if both e and its opposite are stable.

If M is closed, it follows from a well known recurrence theorem of Poincaré that almost all points of E are completely stable. It has been pointed out by E. Hopf ([2], p. 712) that the same result holds if the space of motions under consideration is of finite volume. Since all manifolds of class I have finite area, the volume of the corresponding space of motions is finite and we have the following result.

THEOREM 7.1. *If M is a manifold of class* I, *almost all points of E are completely stable.*

The point e of E will be said to be *unstable* (E. Hopf, fliehende) if, gr being the directed geodesic ray determined by e, and $P_1, P_2, \cdots$ being any sequence of points of gr such that $\lim_{n\to\infty} P_n = A$, A the point at infinity of gr, then the points $P_1, P_2, \cdots$, together with all congruent points, have no cluster point in Ψ. This is equivalent to the condition that the geodesic ray of M determined by e shall eventually pass out of and remain outside of any finite region of M. An element which is not stable is not necessarily unstable. In the case of closed manifolds of class I there are no unstable elements, but not all elements are stable. However, it has been shown by E. Hopf ([2], Theorem 1) that for a class of dynamical systems which include all those under consideration here, *almost all points of E are either stable or unstable.*

The element e will be said to be *transitive* if the elements on the geodesic ray gr determined by e and on the set of geodesic rays congruent to gr form a set which is everywhere dense in E. It is known that there exist transitive elements on manifolds of class I or class II (cf. Hedlund [1]). There are no transitive elements on a manifold of class III.

The property that almost all the elements be transitive can be shown to be equivalent to a simple property of the flow defined by the geodesics (cf. Tuller [1], p. 92). To that end, if S is a set in E, let S_r denote the set of elements on the geodesic rays determined by the points of S. Let S^* be the set consisting of S together with all congruent sets. It is easily shown that the sets $(S_r)^*$ and $(S^*)_r$ are identical and this set will be denoted by S_r^*.

DEFINITION 7.1. *The manifold M has Property* B *if, S being any measurable set of E of positive measure, the set S_r^* is everywhere dense in E.*

THEOREM 7.2. *A necessary and sufficient condition that almost all elements be transitive is that Property* B *hold.*

If almost all points of E are transitive and S is a set of positive measure, S must contain a transitive element. But then the set S_r^* contains all the elements on the ray determined by this element, as well as on all congruent rays, and thus the set S_r^* must be everywhere dense in E. It follows that Property B holds.

Assuming that Property B holds, let $O_1, O_2, \cdots$ be a sequence of open sets in E such that $E \subset \sum_{n=N}^{\infty} O_n$ for all N, and such that the maximum Euclidean diameter of O_n approaches zero as n becomes infinite. Let E_n be the set of elements such that if $e \subset E_n$, the set e_r^* has an element in O_n. Then the set $\{C_E(E_n)\}_r^*$ contains no element of O_n, and since Property B holds, it follows that $C_E(E_n)$ must be a zero set. This implies that the set $\prod_{n=1}^{\infty} E_n$ constitutes almost all points of E. But if an element belongs to all E_n, it is evidently transitive and thus almost all points of E are transitive. The proof of the theorem is complete.

8. **The measure of the transitive elements on manifolds of class** I. Let M be a manifold of class I and let S' be a set of positive measure of E. According to Theorem 7.1, almost all points of E are completely stable, so that, except for a set of measure zero, the points of S' are completely stable. Let S be the set of completely stable points of S'.

Let $S(u, v)$, $u^2 + v^2 < 1$, denote the set of values of ϕ such that (u, v, ϕ) belongs to S. Then, according to a well known theorem of point set theory, one of the sets $S(u_0, v_0)$ is a measurable linear set of positive measure (in the Euclidean sense). Furthermore, the set $S(u_0, v_0)$ must contain a value ϕ_0 at which the linear metric density is unity. Let m be the element (u_0, v_0, ϕ_0). This point can and will be chosen so that $0 < \phi_0 < 2\pi$.

The element m determines a directed geodesic ray with initial point $M_0(u_0, v_0)$. Let B be the point at infinity of this geodesic ray and let this ray be denoted by $gr(M_0, B)$. Since the element m is stable, there exists on $gr(M_0, B)$ a sequence of points M_n $(n = 1, 2, \cdots)$ such that $\lim_{n \to \infty} M_n = B$, and if m_n is the element of $gr(M_0, B)$ at M_n, there is an element e_n congruent to m_n such that $\lim_{n \to \infty} e_n = m$. If $\bar{m}_n$ is the element opposite m_n and $\bar{e}_n$ is the element opposite e_n, it follows that $\lim_{n \to \infty} \bar{e}_n = \bar{m}$, where $\bar{m}$ is the element opposite m. Let A be the point at infinity of the geodesic ray determined by $\bar{m}$. Then if T_n denotes the transformation of the group F taking m_n into e_n, we have $T_n(\bar{m}_n) = \bar{e}_n$, and since $\lim_{n \to \infty} \bar{e}_n = \bar{m}$, it follows that $\lim_{n \to \infty} T_n(M_0) = A$.

Let C_n' be the geodesic circle with center M_0 and passing through M_n, and let C_n be the geodesic circle $T_n(C_n')$. Then, as n becomes infinite, the center $T_n(M_0)$ of C_n approaches A and C_n has on it a point P_n, the point bearing e_n (and $\bar{e}_n$), such that $\lim_{n \to \infty} P_n = M_0$. If Q is an arbitrary point of the horocycle

$C(M_0, A)$, Theorem **5.1** states that there exist a constant L and sequence of points Q_n $(n = 1, 2, \cdots)$ such that $\lim_{n\to\infty} Q_n = Q$ and Q_n lies on the arc of C_n of length L with midpoint P_n. Thus, if we apply the transformation T_n^{-1}, if l_n is the arc of C'_n of length L and with midpoint M_n, l_n has on it a point Q'_n such that $\lim_{n\to\infty} T_n(Q'_n) = Q$.

If the point M_0 is joined to the points of l_n by directed geodesic segments with initial points at M_0, the values of ϕ of the initial elements of these segments form a set $\alpha_n \leqq \phi \leqq \beta_n$, where $\alpha_n < \phi_0 < \beta_n$. Of these values of ϕ, let those which belong to $S(u_0, v_0)$ be denoted by E_n^S and those which do not by F_n^S. Since the length of l_n is L and does not change with n, $\lim_{n\to\infty} | \alpha_n - \beta_n | = 0$. Since the density of the set $S(u_0, v_0)$ is unity at ϕ_0, it follows that

$$\lim_{n\to\infty} \frac{\mu F_n^S}{\mu E_n^S} = 0,$$

where μ denotes linear measure.

Let E_n denote the set of points of l_n determined by the geodesic segments for which the ϕ of the initial element is in E_n^S, and let F_n denote the remaining points of l_n. Then if the measure of sets on l_n in terms of the arc length on l_n is denoted by ν, we have

$$\nu E_n = \int_{E_n^S} \frac{ds}{d\phi}\, d\phi = \int_{E_n^S} G(r, \phi)\, d\phi,$$

$$\nu F_n = \int_{F_n^S} \frac{ds}{d\phi}\, d\phi = \int_{F_n^S} G(r, \phi)\, d\phi,$$

where the integrals are Lebesgue integrals and $G(r, \phi)$ is that determined by setting up geodesic polar coördinates with M_0 as center. If G_M^n denotes the maximum and G_m^n the minimum of $G(r, \phi)$ on l_n, it follows that

$$\frac{\nu F_n}{\nu E_n} = \frac{\int_{F_n^S} G(r, \phi)\, d\phi}{\int_{E_n^S} G(r, \phi)\, d\phi} \leqq \frac{G_M^n \cdot \mu F_n^S}{G_m^n \cdot \mu E_n^S}. \tag{8.1}$$

For n sufficiently large, the radius of the geodesic circle of which l_n is a segment is greater than unity, and it follows from Theorem 4.1 that

$$\frac{G_M^n}{G_m^n} < B.$$

If we combine this with (8.1), there results

$$\lim_{n\to\infty} \frac{\nu F_n}{\nu E_n} \leq B \lim_{n\to\infty} \frac{\mu F_n^S}{\mu E_n^S} = 0. \tag{8.2}$$

Since $\nu E_n + \nu F_n = L$, it follows from (8.2) that

$$\lim_{n\to\infty} \nu E_n = L. \tag{8.3}$$

This implies that the length of the segment of maximum length of l_n which does not contain a point of E_n must approach zero as n becomes infinite. Consequently, there exists a sequence of points Q_n'' $(n = 1, 2, \cdots)$, $Q_n'' \subset E_n$, such that the length of the segment $Q_n'Q_n''$ of l_n approaches zero as n becomes infinite. But then the geodesic distance between Q_n' and Q_n'' must also approach zero, and since $\lim_{n\to\infty} T_n(Q_n') = Q$, it follows that $\lim_{n\to\infty} T_n(Q_n'') = Q$.

Let m_n'' be the element at Q_n'' of the directed geodesic ray with initial point at M_0 and passing through Q_n''. Then by definition of S_r^*, m_n'' and $T_n(m_n'')$ belong to S_r^*. Since $\lim_{n\to\infty} T_n(M_0) = A$ and $\lim_{n\to\infty} T_n(Q_n'') = Q$, it follows that $\lim_{n\to\infty} T_n(m_n'') = q$, where q is the element at Q opposite to the initial element of $gr(Q, A)$. It has been shown by Grant (Grant, Theorem 6.3) that $gr(Q, A)$ is orthogonal to $C(M_0, A)$ at Q, so that if $C_R(M_0, A)$ denotes the right horocycle determined by M_0 and A, the element q is obtained by rotating the element of $C_R(M_0, A)$ at Q through the angle $\frac{1}{2}\pi$.

Also it has been shown by Grant (Grant, Theorem 2.4′) that since the element m is completely stable, and thus $\bar{m}$ is stable, $C_R(M_0, A)$ is transitive; that is, the elements of $C_R(M_0, A)$ together with congruent elements form a set which is everywhere dense in E. But then the elements obtained by rotating each of the elements of $C_R(M_0, A)$ through the angle $\frac{1}{2}\pi$ must form a set such that it, together with the congruent sets, forms a set which is everywhere dense in E. Thus if E_0 is an arbitrary open set of E, the point Q can be so chosen that either q or a congruent element lies in E_0. It has been shown that there exists a sequence of elements $T_n(m_n'')$ of S_r^* such that $\lim_{n\to\infty} T_n(m_n'') = q$. But then, since S_r^* includes all elements congruent to the elements $T_n(m_n'')$, S_r^* must contain elements in E_0. The set S_r^* is everywhere dense in E, and Property B holds. With the aid of Theorem 7.2, the proof of the following theorem is complete.

THEOREM 8.1. *Almost all the elements on a manifold of class* I *are transitive.*

A directed geodesic is *transitive* if the elements on the geodesic and on all congruent geodesics together form a set which is everywhere dense in E. If one element of the geodesic is transitive, the geodesic is transitive. A set of geodesics will be said to constitute *almost all the geodesics* if the elements on these and congruent geodesics constitute almost all points of E. The preceding theorem can now be restated as follows.

THEOREM 8.2. *Almost all geodesics on a manifold of class* I *are transitive.*

In view of the remarks in §6 concerning closed surfaces of negative curvature, we can state the following result.

COROLLARY 8.2′. *Almost all geodesics on a closed orientable surface of class* C^4 *and of negative curvature are transitive.*

For if the functions defining the surface are of class C^8, the coefficients of the first fundamental quadratic form are of class C^7, and the surface is included among the manifolds of class I.

COROLLARY 8.2″. *Almost all the geodesics on a closed non-orientable surface of class* C^8 *and of negative curvature are transitive.*

For such a surface has a closed orientable covering surface of multiplicity two and a transitive geodesic on the covering surface has as correspondent a transitive geodesic in the non-orientable surface. The statement of the corollary is thus a consequence of Corollary 8.2′.

9. **The measure of the unstable elements on manifolds of class** III. Let M be a manifold of class III, that is, one for which the defining Fuchsian group F is of the second kind. Then the fundamental region R_0 abuts on the circle U along one or more arcs, and the region R, which consists of R_0 together with its reflection in U, contains the interior points of these arcs in its interior (cf. Ford [1], p. 74). Let α denote the set of interior points of these arcs.

Suppose that A is a point of the set α and e is an element determining a geodesic ray with point at infinity A. We show that e is unstable. Recall that e has been defined as unstable if, $P_1, P_2, \cdots$ being any sequence of points of the geodesic ray determined by e such that $\lim_{n\to\infty} P_n = A$, these points together with all congruent points have no cluster point in Ψ. Since A is an interior point of R, for n sufficiently large, the point P_n lies in R_0. But then (Ford [1], Theorem 9, p. 71) P_n is nearer the center of U than any point congruent to it. Since $\lim_{n\to\infty} P_n = A$, and any closed region lying within U is covered by a finite number of transforms of R_0 (Ford [1], p. 70), it follows that points congruent to the set P_n $(n = 1, 2, \cdots)$ cannot have a cluster point in Ψ. The following lemma has been proved.

LEMMA 9.1. *If A is a point of U belonging to the set α, and e is an element determining a geodesic ray with point at infinity A, e is unstable.*

THEOREM 9.1. *Almost all elements on manifolds of class* III *are unstable.*

Suppose that the statement of the theorem is not true. Then it follows from a theorem of Hopf (cf. E. Hopf [2], Theorem 1) that there must be a set S of stable elements of E such that S is of positive measure. Let the element $m(u_0, v_0, \phi_0)$ be chosen as in the proof of Theorem 8.1, where the present set S takes the place of the set S used in the proof of Theorem 8.1. Let M_0 again be the point (u_0, v_0) and let A be the point at infinity of the geodesic ray with initial element $\bar{m}$, the element opposite m. If q is an element obtained by rotating an arbitrary element of $C_R(M_0, A)$ through the angle $\frac{1}{2}\pi$, the proof of Theorem 8.1 shows that in the case under consideration the set S_r^* contains a sequence of elements q_n $(n = 1, 2, \cdots)$ such that $\lim_{n\to\infty} q_n = q$.

Now let $D \neq A$ be a point belonging to the set α. The geodesic with initial

point at infinity A and terminal point at infinity D cuts across $C_R(M_0, A)$ orthogonally at a point Q. Choosing this as the point Q above, if D_n denotes the point at infinity of the geodesic ray with initial element q_n, since $\lim_{n \to \infty} q_n = q$, we see that $\lim_{n \to \infty} D_n = D$. But then for n sufficiently large, D_n belongs to the set α, and it follows from Lemma 9.1 that q_n, n sufficiently large, must be unstable. It is easily shown that if an element is stable, the same is true of all the elements on the geodesic ray determined by this element. Since it was assumed that all elements of S were stable, the same would be true of all elements in S_r^*. But if q_n is unstable, it cannot be stable, and from this contradiction we infer the truth of the theorem.

Bibliography

L. R. Ford.
1. *Automorphic Functions*, New York, 1929.

J. Hadamard.
1. *Les surfaces à courbures opposées et leur lignes géodésiques*, Journal de Mathématiques pures et appliquées, (5), vol. 4(1898), pp. 27–73.

G. A. Hedlund.
1. *Two-dimensional manifolds and transitivity*, Annals of Mathematics, vol. 37(1936), pp. 534–542.
2. *Fuchsian groups and transitive horocycles*, this Journal, vol. 2(1936), pp. 530–542.

E. Hopf.
1. *Fuchsian groups and ergodic theory*, Transactions of the American Mathematical Society, vol. 39(1936), p. 299.
2. *Zwei Sätze über den wahrscheinlichen Verlauf der Bewegungen dynamischer Systeme*, Mathematische Annalen, vol. 103(1930), pp. 710–719.

M. Morse.
1. *Instability and transitivity*, Journal de Mathématiques pures et appliquées, (9), vol. 14(1935), pp. 49–71.

A. Tuller.
1. *The measure of the transitive geodesics on certain three-dimensional manifolds*, this Journal, vol. 4(1938), pp. 78–94.

Bryn Mawr College.

A PROOF THAT EVERY UNIFORMLY CONVEX SPACE IS REFLEXIVE

By B. J. Pettis

The purpose of the present note is to communicate an independent proof of a result of Milman's [6][1] to the effect that every uniformly convex space is necessarily reflexive. Our proof is quite different from Milman's, being based on the use of bounded additive measure functions rather than on that of transfinite closure for closed convex sets.

Let $\mathfrak{X} = [x]$ be a Banach space, $\bar{\mathfrak{X}} = [\gamma]$ its adjoint, and $\bar{\bar{\mathfrak{X}}} = [F]$ the adjoint of $\bar{\mathfrak{X}}$. The space $\mathfrak{X}$ is said to be *reflexive*[2] if for each $F_0 \epsilon \bar{\bar{\mathfrak{X}}}$ there is an $x_0 \epsilon \mathfrak{X}$ such that $F_0(\gamma) = \gamma(x_0)$ holds for all $\gamma \epsilon \bar{\mathfrak{X}}$. The concept of $\mathfrak{X}$ being a *uniformly convex* space, a concept of Clarkson's [2], may be defined in the following fashion: *given $\epsilon > 0$ there is a $\zeta_\epsilon > 0$ with the property that*

$$(*)\qquad \text{if } x, y \epsilon \mathfrak{X} \text{ have } \|x\| = \|y\| = 1 \text{ and if } \|x - y\| \geqq \epsilon, \text{ then } \|x + y\| \leqq 2 - \zeta_\epsilon.$$

The theorem may now be stated as follows.

Theorem (Milman). *If $\mathfrak{X}$ is isomorphic to a uniformly convex space, then $\mathfrak{X}$ is reflexive.*

We first establish two lemmas.

Lemma 1.[3] *If $\mathfrak{X}$ is uniformly convex, then given $\gamma_0 \epsilon \bar{\mathfrak{X}}$ with $\|\gamma_0\| \neq 0$, there exists a unique $x_0 \epsilon \mathfrak{X}$ satisfying the conditions $\|x_0\| = 1$ and $\gamma_0(x_0) = \|\gamma_0\|$. Moreover, given $\epsilon > 0$, there exists a $\delta_\epsilon > 0$ such that if x and y in $\mathfrak{X}$ and γ in $\bar{\mathfrak{X}}$ satisfy the conditions $\|x\| = 1$, $\|y\| \leqq 1$, $\|x - y\| \geqq \epsilon$, and $\gamma(x) = \|\gamma\|$, then $\gamma(y) \leqq (1 - \delta_\epsilon)\|\gamma\|$.*

In proving the first part it is clearly sufficient to consider only the case $\|\gamma_0\| = 1$. By definition of $\|\gamma_0\|$ there then exists a sequence $\{x_n\}$ with $\|x_n\| = 1$ and $1 \geqq \gamma_0(x_n) > 1 - n^{-1}$. To see that $\{x_n\}$ is a Cauchy sequence

This note, as first submitted to the editors, was part of the manuscript of a paper entitled *On differentiation in Banach spaces* which was presented to the American Mathematical Society, September 6, 1938, and abstracted in the Bull. Amer. Math. Soc., vol. 44, no. 7 (July, 1938), p. 486. The manuscript was received August 30, 1938, before receipt of Milman's article in this country. At the suggestion of the referee the present paper is being published separately. It was received in the above form April 8, 1939.

[1] Numbers in brackets refer to the list of references.

[2] Such spaces were introduced by Hahn [4] under the name of *regular*. The present term *reflexive* is due to Lorch.

[3] The first statement contained in Lemma 1 was discovered independently by J. A. Clarkson and E. J. McShane in 1936. I am grateful to them for permission to include it here.

let $\epsilon > 0$ be given, let ζ_ϵ correspond to ϵ according to (*), and choose $n_\epsilon > 2/\zeta_\epsilon$. Then $m \geqq n_\epsilon$, $n \geqq n_\epsilon$ imply that

$$\gamma_0(x_m + x_n) = \gamma_0(x_m) + \gamma_0(x_n) > 2 - \frac{2}{n_\epsilon} > 2 - \zeta_\epsilon$$

and hence that $\| x_m + x_n \| > 2 - \zeta_\epsilon$. Since $\| x_m \| = \| x_n \| = 1$, this last inequality yields $\| x_m - x_n \| < \epsilon$ whenever $m, n \geqq n_\epsilon$, since otherwise (*) is contradicted.

Let $x_0 = \lim_n x_n$. Then

$$\| x_0 \| = \lim_n \| x_n \| = 1, \qquad \gamma_0(x_0) = \lim_n \gamma_0(x_n) = \| \gamma_0 \|.$$

If there were another point x_1 such that $\| x_1 \| = 1$ and $\gamma_0(x_1) = 1$, and if $\| x_1 - x_0 \| = \rho > 0$, then from (*) it would follow that $\| x_0 + x_1 \| \leqq 2 - \zeta_\rho < 2$ and hence

$$2 > | \gamma_0(x_0 + x_1) | = \gamma_0(x_0) + \gamma_0(x_1) = 2,$$

an impossibility. This establishes the first half of the lemma.

To prove the second part let any $\epsilon > 0$ be given and set $\eta = \min(\frac{1}{2}, \frac{1}{2}\epsilon)$. It follows that if ζ_η corresponds to η according to (*), then $\delta_\epsilon \equiv \min(\zeta_\eta, \eta)$ is a satisfactory δ_ϵ. For if $\| y \| \leqq 1 - \eta$, then $\gamma(y) \leqq \| \gamma \| \cdot (1 - \eta) \leqq \| \gamma \| \cdot (1 - \delta_\epsilon)$. And if $1 \geqq \| y \| \geqq 1 - \eta$, let $z = y / \| y \|$. Should $\| x - y \| \geqq \epsilon$ hold, we must have

$$\| x - z \| \geqq \| x - y \| - \| y - z \| \geqq \epsilon - \| y \| \cdot \left| 1 - \frac{1}{\| y \|} \right| = \epsilon - (1 - \| y \|)$$

$$\geqq \epsilon - \eta \geqq \eta,$$

where $\| x \| = \| z \| = 1$. From (*) it then follows that $\| x + z \| \leqq 2 - \zeta_\eta$, and therefore when $\gamma(x) = \| \gamma \|$, it is true that

$$\gamma(z) = \gamma(z + x) - \gamma(x) \leqq (2 - \zeta_\eta) \cdot \| \gamma \| - \| \gamma \|$$

$$= (1 - \zeta_\eta) \cdot \| \gamma \| \leqq (1 - \delta_\epsilon) \cdot \| \gamma \|.$$

Thus $\gamma(y) = \| y \| \cdot \gamma(z) \leqq \| \gamma \| \cdot (1 - \delta_\epsilon)$. The inequality $\gamma(y) \leqq \| \gamma \| \cdot (1 - \delta_\epsilon)$ then holds uniformly for all x, y, γ satisfying the conditions $\| x \| = 1$, $\| y \| \leqq 1$, $\| x - y \| \geqq \epsilon$, and $\gamma(x) = \| \gamma \|$.

The second lemma is a sharpening of a result of Goldstine's [3].

Lemma 2. *If $\mathfrak{X}$ is an arbitrary B-space and $F_0(\gamma)$ is any linear functional defined over $\overline{\mathfrak{X}} = [\gamma]$, then there exists a function $\beta(E)$ having the properties:*

(i) *$\beta(E)$ is defined, bounded, and additive over all subsets E of the unit sphere S of $\mathfrak{X}$;*

(ii) *$\beta(E)$ is non-negative;*

(iii) $\| F_0 \| = \| \beta \| = \mathrm{Var}\,(\beta; S)$;

(iv) $F_0(\gamma) = \int_S \gamma(x)\, d\beta,\ \gamma \in \overline{\mathfrak{X}}$,

the integral in (iv) *being the Radon-Stieltjes.*[4]

Given F_0, Goldstine has proved the existence of an $\alpha(E)$ satisfying (i), (iii), and (iv). Let $\alpha(E) = \pi(E) - \nu(E)$ be the Jordan decomposition of α, where π and ν have properties (i) and (ii). For each set E in S let $P(E)$ be the projection of E through the origin, i.e., $P(E)$ is the set of all x in S having $-x$ in E, and define $\nu_0(E) = \nu(P(E))$. Then $\nu_0(E)$ also has properties (i) and (ii), and $\nu_0(S) = \nu(S)$. If $\beta(E) \equiv \pi(E) + \nu_0(E)$, it now follows that $\beta(E)$ has properties (i), (ii), and (iii); the first two are evident and the third follows from the fact that

$$\| F_0 \| = \mathrm{Var}\,(\alpha; S) = \pi(S) + \nu(S) = \beta(S) = \mathrm{Var}\,(\beta; S),$$

the last equality arising from β being non-negative.

For every $\gamma \in \overline{\mathfrak{X}}$ we now obtain from the definition of the integral that

$$\int_S -\gamma(x)\, d\nu = \int_S \gamma(-x)\, d\nu = \int_S \gamma(x')\, d\nu_0,$$

i.e., we make the "change of variable" $x' = -x$. From property (iv) of $\alpha(E)$,

$$F_0(\gamma) = \int_S \gamma(x)\, d\pi - \int_S \gamma(x)\, d\nu = \int_S \gamma(x)\, d\pi + \int_S -\gamma(x)\, d\nu$$

$$= \int_S \gamma(x)\, d\pi + \int_S \gamma(x)\, d\nu_0 = \int_S \gamma(x)\, d\beta.$$

This completes the proof of Lemma 2.

In proving the main theorem we may assume $\mathfrak{X}$ itself to be uniformly convex, since any space isomorphic to a reflexive space is necessarily reflexive.

We thus suppose $\mathfrak{X}$ to be uniformly convex, and seek to show that for each $F_0 \in \bar{\mathfrak{X}}$ there is an $x_0 \in \mathfrak{X}$ such that $F_0(\gamma) = \gamma(x_0)$ for all $\gamma \in \overline{\mathfrak{X}}$. It is sufficient to consider only those points in $\mathfrak{X}$ having unit norm. For such an F_0 there must be elements γ_n $(n = 1, 2, \cdots)$ in $\overline{\mathfrak{X}}$ such that

(1) $$\| \gamma_n \| = 1 \qquad \text{and} \qquad 1 = \| F_0 \| = F_0(\gamma_n) \geq 1 - \frac{1}{n}.$$

By Lemma 1 there exists a sequence $\{x_n\}$ satisfying the conditions

(2) $$\| x_n \| = 1 \quad \text{and} \quad \gamma_n(x_n) = \| \gamma_n \| = 1.$$

It will be proved first that $\{x_n\}$ is a Cauchy sequence and then that $x_0 = \lim_n x_n$ has the property that $F_0(\gamma) = \gamma(x_0)$ for all γ.

For $F_0(\gamma)$ consider the $\beta(E)$ of Lemma 2 which has the properties

(3) $$1 = \| F_0 \| = \mathrm{Var}\,(\beta; S) = \beta(S),$$

[4] For a description of this integral see [5].

$$(4) \qquad F_0(\gamma) = \int_S \gamma(x)\, d\beta, \qquad \gamma \in \bar{\bar{\mathfrak{X}}}.$$

If we use (4), (1) leads to the inequality

$$(5) \qquad 1 - \frac{1}{n} < \int_S \gamma_n(x)\, d\beta = \int_{S_{n,\epsilon}} \gamma_n(x)\, d\beta + \int_{S - S_{n,\epsilon}} \gamma_n(x)\, d\beta,$$

where $S_{n,\epsilon} = S[\| x - x_n \| < \epsilon]$. In view of (2) and Lemma 1, given $\epsilon > 0$ there exists a positive δ_ϵ that is independent of x_n and has the property that $1 - \delta_\epsilon > \gamma_n(x)$ holds for all x in $S - S_{n,\epsilon}$. Since this last inequality holds and $\beta(E)$ is non-negative, (5) gives us

$$1 - \frac{1}{n} < \int_{S_{n,\epsilon}} \gamma_n(x)\, d\beta + (1 - \delta_\epsilon)\beta(S - S_{n,\epsilon}),$$

or

$$(5') \qquad 1 - \frac{1}{n} < \beta(S_{n,\epsilon}) + (1 - \delta_\epsilon)\beta(S - S_{n,\epsilon}),$$

since (2) implies that $| \gamma_n(x) | \leqq \| \gamma_n \| = 1$ holds over S. The function $\beta(E)$ being additive and $\beta(S)$ having the value 1, it follows from (5′) that

$$(6) \qquad \beta(S - S_{n,\epsilon}) < \frac{1}{n \cdot \delta_\epsilon} \qquad (n = 1, 2, \cdots).$$

This is the basic inequality of the proof.

Choosing n_ϵ large enough to satisfy the condition $2 < n_\epsilon \delta_\epsilon$, we can now assert, thanks to (6), that if $m \geqq n_\epsilon$ and $n \geqq n_\epsilon$, then $S_{m,\epsilon} \cdot S_{n,\epsilon}$ is not the vacuous set Λ. For then $\beta(S) = 1$, $\beta(S - S_{m,\epsilon}) < \frac{1}{2}$ and $\beta(S - S_{n,\epsilon}) < \frac{1}{2}$ and these statements imply that

$$\begin{aligned} \beta(S_{n,\epsilon} \cdot S_{m,\epsilon}) &= \beta(S - ((S - S_{n,\epsilon}) + (S - S_{m,\epsilon}))) \\ &= \beta(S) - \beta((S - S_{n,\epsilon}) + (S - S_{m,\epsilon})) \\ &\geqq \beta(S) - (\beta(S - S_{n,\epsilon}) + \beta(S - S_{m,\epsilon})) > 0. \end{aligned}$$

Hence $S_{n,\epsilon} \cdot S_{m,\epsilon} \neq \Lambda$. But it can be inferred from the last statement that for n and $m > n_\epsilon$, $\| x_m - x_n \| < 2\epsilon$. Thus $\{x_m\}$ is a Cauchy sequence.

The final assertion is that $x_0 = \lim_n x_n$ has the desired property. To justify this we use the fact that if $S_{0,\epsilon} = S[\| x - x_0 \| < \epsilon]$, then

$$(7) \qquad \beta(S - S_{0,\epsilon}) = 0 \qquad \text{for every } \epsilon > 0.$$

If $\| x_n - x_0 \| < \xi = \frac{1}{2}\epsilon$, then $S_{0,\epsilon} \supset S_{n,\xi}$ and hence $S - S_{0,\epsilon} \subset S - S_{n,\xi}$. This result and (6) imply that for $n \geqq n_\xi$,

$$0 \leqq \beta(S - S_{0,\epsilon}) \leqq \beta(S - S_{n,\xi}) \leqq \frac{1}{n \cdot \delta_\xi}.$$

Hence (7) holds.

Now consider $F_0(\gamma) - \gamma(x_0)$; in view of (4) and (3) we can write

$$F_0(\gamma) - \gamma(x_0) = \int_S \gamma(x)\, d\beta - \int_S \gamma(x_0)\, d\beta;$$

hence, $\beta(E)$ being non-negative and (7) being true, it follows that

$$|F_0(\gamma) - \gamma(x_0)| \leqq \int_S |\gamma(x) - \gamma(x_0)|\, d\beta = \int_{S_{0,\epsilon}} |\gamma(x) - \gamma(x_0)|\, d\beta$$

$$< \|\gamma\| \cdot \int_{S_{0,\epsilon}} \|x - x_0\|\, d\beta < \|\gamma\| \cdot \epsilon \cdot \beta(S_{0,\epsilon}) = \|\gamma\| \cdot \epsilon$$

is true for any $\epsilon > 0$. Thus $F_0(\gamma) = \gamma(x_0)$ for each γ, and the proof is finished.

References

1. S. Banach, *Théorie des Opérations Linéaires*, Warsaw, 1932.
2. J. A. Clarkson, *Uniformly convex spaces*, Trans. Amer. Math. Soc., vol. 40(1936), pp. 396–414.
3. H. Goldstine, *Weakly complete Banach spaces*, this Journal, vol. 4(1938), pp. 125–131.
4. H. Hahn, *Über lineare Gleichungssysteme in linearen Räumen*, Jour. f. d. reine u. ang. Math., vol. 157(1927), pp. 214–229.
5. T. H. Hildebrandt, *On bounded linear functional operators*, Trans. Amer. Math. Soc., vol. 36(1934), pp. 868–875.
6. D. Milman, *On some criteria for the regularity of spaces of the type* (B), Comptes Rendus de l'Acad. des Sc. de l'URSS, vol. 20(1938), pp. 243–246.

University of Virginia.

DIFFERENTIATION IN BANACH SPACES

By B. J. Pettis

Introduction. Consider a figure[1] R_0 in a Euclidean n-space. According to a classical theorem of Lebesgue, if $\mathfrak{X}$ is the space of real numbers, then every ABV (additive, with bounded variation) function defined to $\mathfrak{X}$ from the figures in R_0 is necessarily differentiable a.e.[2] in R_0. But, as Bochner first pointed out [4], this theorem does not hold for general Banach spaces $\mathfrak{X}$. There exist spaces $\mathfrak{X}$ to which ABV functions may be defined that are differentiable at no point in R_0. Several authors [3, 6, 7, 10, 12, 15] have as a consequence considered the problem of finding conditions on $\mathfrak{X}$ sufficient that every ABV function defined to $\mathfrak{X}$ be differentiable a.e. Here, however, we wish to adopt a somewhat different viewpoint, at least throughout §§1 and 2, the sections fundamental to our discussion: in the principal theorems of the paper, given in §2, the emphasis has been placed on the individual function X_R rather than on the space $\mathfrak{X}$ and the class of all ABV functions having their values in $\mathfrak{X}$. Thus (to put it more explicitly) the conclusions reached in Theorems 2.5, 2.7, 2.8, and 2.9 state, with no restriction on $\mathfrak{X}$, that if a fixed ABV function X_R defined to $\mathfrak{X}$ has a generalized "weak" derivative according to any one of several definitions, then X_R is differentiable a.e.; that is, X_R has a "strong" derivative. In each of these four theorems it is shown that a set of necessary conditions, expressed in terms of linear functionals and apparently quite feeble, are actually of sufficient strength to insure differentiability a.e.

The possible use of these results is not confined to testing the strong differentiability of an individual function having its values in an unrestricted (and perhaps unsatisfactory) space; the theorems can also be applied to the problem considered in the papers cited above, namely, that of testing whether or not a given condition which the space $\mathfrak{X}$ is assumed to satisfy is strong enough to insure the differentiability a.e. of every ABV function defined to $\mathfrak{X}$. The results concerning differentiation that have been obtained in [3], [7], [10], [12], and [15] are here derived in §§3–5 from the theorems of the present §2; in each proof the essential idea is to show that if $\mathfrak{X}$ satisfies the particular condition under consideration, then $\mathfrak{X}$ is weakly compact in one generalized sense or another.

Following §1, in which the necessary definitions have been grouped, the principal theorems will be found in §2. Those dealing with differentiation we

Received August 30, 1938; presented to the American Mathematical Society, September 6, 1938.

[1] Terms used, but undefined, in the present paper will be found in either [18] or [1] (numbers in brackets refer to the list of references at the end). It is supposed that the reader is somewhat familiar with these two treatises.

[2] The phrase "almost everywhere" will be abbreviated throughout to "a.e."

believe to be new. In §3 a theorem of Gelfand's given in [12] and [15] is obtained in slightly improved form (Theorem 3.1), R_0 now being n-dimensional instead of linear and $\mathfrak{X}$ being supposed weakly compact[3] instead of reflexive.[4] Spaces with bases [10] are involved in §4. In §5 there is an extension of a theorem of Gelfand [12] concerning differentiation in spaces subjected to certain separability hypotheses; the extension can be applied to the arbitrary case. Some comments on the differentiation of abstract integrals occupy §6, and two concluding remarks, one on real-valued BV convex functions and the other on a certain assumption made in §2, compose the final section.

Concerning notation, real-valued functions will usually be represented by Greek letters while those defined to more general B-spaces will be put in italics. Functions of points will be in lower case (x_s, y_s, etc.) and functions of figures in capitals (X_R, Y_R, etc.). Among the conventions followed in [1] to these two we shall especially adhere: (i) θ is the zero element of the particular B-space under discussion, and (ii) the adjoint spaces of $\mathfrak{X}$, $\mathfrak{Y}$, $\cdots$ are denoted by $\bar{\mathfrak{X}}$, $\bar{\mathfrak{Y}}$, $\cdots$, respectively, and the adjoint spaces of $\bar{\mathfrak{X}}$, $\bar{\mathfrak{Y}}$, $\cdots$ by $\bar{\bar{\mathfrak{X}}}$, $\bar{\bar{\mathfrak{Y}}}$, $\cdots$, respectively. The function x_s which is identically θ will be written as θ_s.

1. **Preliminary considerations.** In Euclidean n-space let R_0 be a fixed figure and R any figure contained in R_0. A finite number $R_1, \cdots, R_k$ of figures lying in R are *non-overlapping* if $R_i \cdot R_j$ has a vacuous interior whenever $i \neq j$; if it is also true that $\sum_{i=1}^{k} R_i = R$, then these figures form a *partition of* R. Let X_R be a function having the figures R in R_0 for its domain and a Banach space (B-space) $\mathfrak{X}$ as its range. Then X_R is said to be *of bounded variation*, or BV, if $\text{l.u.b.} \sum_{i=1}^{k} \| X_{R_i} \| < \infty$ as the partition $\pi = [R_1, \cdots, R_k]$ ranges over all partitions of R_0. When X_R is BV, the auxiliary real-valued function

$$\text{Var}\,(X_R; R) \equiv \text{l.u.b.} \sum_{i=1}^{k} \| X_{R_i} \|, \qquad \pi \text{ a partition of } R,$$

is defined over the figures R in R_0, and is also BV. A function X_R is termed *additive* if $X_{R_1+R_2} = X_{R_1} + X_{R_2}$ whenever R_1 and R_2 are non-overlapping, and ABV when it is both additive and BV. It is clear that an additive function is always *convex*; that is, X_R satisfies the inequality $\| X_{R_1+R_2} \| \leqq \| X_{R_1} \| + \| X_{R_2} \|$ for all non-overlapping R_1 and R_2.

If s is a fixed point in R_0 and if $\lim X_I / | I |$ exists as I ranges over all the non-degenerate cubes in R_0 that contain s and as $| I |$, the measure[5] of I, tends

[3] For the definition of weak compactness used here see the beginning of §3; this definition is Banach's ([1], p. 239) with weak completeness added.

[4] The definition of a reflexive, or regular [13], space can be found in [12], [15], or [17]. Since reflexiveness implies weak compactness ([11] or [15]), the latter property is a formal weakening of the former. Whether or not it is an actual weakening is, as far as the author knows, still an open question.

[5] Since only measurable sets will be involved in any of the proofs, the adjective "measurable" as applied to sets will usually be omitted.

to 0, then X_R is said to be *differentiable at s*. When X_R is differentiable at s, there exists in $\mathfrak{X}$, since $\mathfrak{X}$ is complete, a unique element x_s such that $x_s = \lim X_I / |I|$; this element is *the derivative of X_R at s*, and X_R is *differentiable to the value x_s at s*. The particular term *singular* is applied to an ABV function X_R that a.e. in R_0 is differentiable to the zero element θ of $\mathfrak{X}$; i.e., X_R is singular if it is ABV and if a.e. in R_0 it is true that $\lim \|X_I\| / |I| = 0$ when I is subject to the conditions above. In proofs involving either additive functions or convex functions we shall often find the following elementary theorem useful:

If X_R is convex and BV, then

(I) *the real-valued function $\Omega_R \equiv \mathrm{Var}\,(X_R\,; R)$ is ABV;*

(II) *the difference quotients of X_R are bounded at almost every point of R_0;*

(III) *if Ω_R is singular, X_R is differentiable to θ a.e.*

Conclusion I is a direct consequence of X_R being convex and BV; the proof is left to the reader. Conclusion III follows immediately from the inequality $0 \leqq \|X_R\| \leqq \Omega_R$, and II results[6] from combining this inequality with the fact that the real-valued function Ω_R, being ABV, must have its difference quotients bounded a.e.

We now consider various definitions of generalized or weak derivatives. Let X_R be defined to $\mathfrak{X}$ from the figures in R_0 and let x_s be a function defined a.e. in R_0 and having its range in $\mathfrak{X}$. If $\mathfrak{Z} = [\zeta]$ is a set in $\bar{\mathfrak{X}}$, the adjoint of $\mathfrak{X}$, then X_R is said to be *$\mathfrak{Z}$-differentiable to x_s* if there exists a measurable set S in R_0 such that (i) $|S| = |R_0|$ and (ii) $s \in S$ implies that for every $\zeta \in \mathfrak{Z}$ the real-valued figure function $\zeta(X_R)$ is differentiable at s to the value $\zeta(x_s)$; under these circumstances x_s is said to be a *$\mathfrak{Z}$-derivative of X_R*. If X_R has at least one $\mathfrak{Z}$-derivative x_s, and if any other $\mathfrak{Z}$-derivative y_s of X_R is necessarily equivalent to x_s, that is, the equality $x_s = y_s$ must hold a.e. in R_0, then X_R will be said to have a *unique $\mathfrak{Z}$-derivative.* We note that if the set $\mathfrak{Z}$ of linear functionals forms a total set over $\mathfrak{X}$, and if X_R has a $\mathfrak{Z}$-derivative, then this derivative is unique. In particular, if X_R has an $\bar{\mathfrak{X}}$-derivative x_s, then x_s is unique; X_R is then said to be *weakly differentiable a.e.* and x_s is the *weak derivative of X_R*.

More generally, we say that X_R is *$\mathfrak{Z}$-pseudo-differentiable to x_s* (and x_s is a *$\mathfrak{Z}$-pseudo-derivative of X_R*) if [16] for each $\zeta \in \mathfrak{Z}$ the function $\zeta(X_R)$ is differentiable a.e. to $\zeta(x_s)$. If X_R has at least one $\mathfrak{Z}$-pseudo-derivative, and if any two such derivatives are necessarily equivalent, then X_R is said to have a *unique $\mathfrak{Z}$-pseudo-derivative.* It is evident that any $\mathfrak{Z}$-derivative is a $\mathfrak{Z}$-pseudo-derivative, and that for denumerable $\mathfrak{Z}$ the converse is true. Thus if $\mathfrak{Z}$ is both denumerable and total and x_s is a $\mathfrak{Z}$-pseudo-derivative of X_R, then x_s is a $\mathfrak{Z}$-derivative of X_R and this derivative is unique.

By the *span* of a set $\mathfrak{G}$ in $\mathfrak{X}$ is meant the smallest closed linear subspace $\mathfrak{Y}$ in $\mathfrak{X}$ that contains $\mathfrak{G}$. If $\pi_n = [R_{n,1}, \cdots, R_{n,k_n}]$ $(n = 1, 2, \cdots)$ is a sequence of partitions of R_0 and X_R is a fixed ABV function, then the *X_R-span of the par-*

[6] A proof that II holds for BV functions whether convex or not is contained in [7], p. 409.

titions $\{\pi_n\}$ is the span of the denumerable set $\{X_{R_{n,i}}\}$. The *span of* x_s (or of X_R) is the span of the set of functional values of x_s (or of X_R); if this span is separable, then x_s (or X_R) is said to be *separably-valued*. A function x_s that is equivalent to a separably-valued function will be referred to as *essentially separably-valued*.

The two final definitions, which are fundamental for our purposes, are these:

DEFINITION 1.1. *If* $\{\gamma_n\}$ *is a sequence in* $\bar{\mathfrak{X}}$ *and* $\mathfrak{Y}$ *is a set in* $\mathfrak{X}$, *then* $\{\gamma_n\}$ *is said to have property* $N(\mathfrak{Y})$ *if* $\| \gamma_n \| \leqq 1$ $(n = 1, 2, \cdots)$ *and* $\| y \| = \limsup_n |\gamma_n(y)|$ *for every* $y \in \mathfrak{Y}$.

DEFINITION 1.2. *Given* X_R, *an* X_R-*maximal sequence is a sequence* $\pi_n = [R_{n,1}, \cdots, R_{n,k_n}]$ $(n = 1, 2, \cdots)$ *of partitions of* R_0 *such that* $\lim_{n\to\infty} \sum_{i=1}^{k_n} \| X_{R_{n,i}} \| = \text{Var}\,(X_R\,;\,R_0)$.

Concerning Definition 1.2 it is obvious that X_R-maximal sequences always exist for any X_R. Likewise, if a set $\mathfrak{Y}$ is contained in a separable closed linear subspace of $\mathfrak{X}$, i.e., if the span $\mathfrak{Z}$ of $\mathfrak{Y}$ is separable, then there exists at least one sequence in $\bar{\mathfrak{X}}$ that has property $N(\mathfrak{Y})$; it is sufficient to take a sequence $\{\gamma_n'\}$ weakly dense (as functionals) in the unit sphere of $\bar{\mathfrak{Z}}$ ([1], p. 124, Theorem 4) and then extend each γ_n' to form an element γ_n of $\bar{\mathfrak{X}}$ while preserving the norm. It is to be noted that from the last condition in Definition 1.1 it follows that every sequence $\{\gamma_n\}$ having property $N(\mathfrak{Y})$ forms a total set of functionals over $\mathfrak{Y}$.

2. **Necessary conditions that are also sufficient in order that an ABV function X_R be differentiable a.e.** A function x_s is a *simple function* if it is constant on each of a finite number of measurable sets whose sum is R_0; according to Bochner [5], x_s is *measurable* if it is the limit a.e. of a sequence of simple functions. In proving the fundamental result of the present paper, two preliminary theorems concerning measurability and integrability of abstract functions will be used. The first, which gives a condition sufficient that x_s be measurable, is an extension of a known theorem [12, 16].

THEOREM 2.1. *Let* x_s *have a separable span* $\mathfrak{Y}$. *If* $\gamma_m(x_s)$ *is measurable for every* γ_m *in some sequence* $\{\gamma_m\}$ *that has property* $N(\mathfrak{Y})$, *then* x_s *is measurable.*

Since $\{\gamma_m\}$ has property $N(\mathfrak{Y})$, where $\mathfrak{Y}$ is the span of x_s, it follows that for every s we have $\| x_s \| = \limsup_m |\gamma_m(x_s)|$, so that $\| x_s \|$ is measurable, being the lim sup of measurable functions. About each of the points x_s in the span $\mathfrak{Y}$ put an open sphere of radius n^{-1}. The space $\mathfrak{Y}$ being metric and separable, a denumerable number $\mathfrak{K}^{ni}$ $(i = 1, 2, \cdots)$ of these spheres together cover all the functional values of x_s. If x^{ni} is the center of $\mathfrak{K}^{ni}$, then $x^{ni} \in \mathfrak{Y}$ and hence $\mathfrak{Y}$ is the span of $x_s - x^{ni}$; moreover, $\gamma_m(x_s - x^{ni}) = \gamma_m(x_s) - \gamma_m(x^{ni})$ is clearly measurable for every i and m. Hence an argument used above shows that

$\| x_s - x^{ni} \|$ is measurable for each i. This implies that $G^{ni} = R_0[\| x_s - x^{ni} \| < n^{-1}]$ is a measurable set in R_0, and since every point x_s is in some $\mathfrak{K}^{ni}$, it follows that $\sum_{i=1}^{\infty} G^{ni} = R_0$. Hence if $H^{nj} = G^{nj} - \sum_{i=1}^{j-1} G^{ni}$, then H^{nj} $(j = 1, 2, \cdots)$ is a sequence of disjoint measurable sets whose sum is R_0. Now define x_s^n to have the constant value x^{nj} on the set H^{nj}; this function is clearly measurable and satisfies the condition $\| x_s - x_s^n \| < n^{-1}$ for all s, since $\sum_j H^{nj} = R_0$ and $s \in H^{nj}$ implies that $\| x_s - x^{nj} \| < n^{-1}$. Thus x_s can be uniformly approximated in R_0 by measurable functions and therefore is itself measurable.

The abstract integral defined by Bochner [5] plays an important rôle in the theory of B-space differentiation [6, 7, 10, 12, 15], since this integral is ABV and a.e. differentiable [5] and every derivative (when it exists) of an ABV function is integrable [7, Theorem 5], that is, integrable under Bochner's definition. Recalling that a necessary and sufficient condition that x_s be integrable is that it be measurable and that $\| x_s \|$ be summable in the Lebesgue sense, we shall find in the next theorem (which is a generalization of Theorem 5 of [15]) a condition sufficient for the integrability of a pseudo-derivative.

THEOREM 2.2. *Suppose x_s has a separable span $\mathfrak{Y}$. If $\mathfrak{Z} = \{\gamma_j\}$ has property $N(\mathfrak{Y})$, and if x_s is a $\mathfrak{Z}$-pseudo-derivative of some function X_R, then x_s is measurable. If in addition X_R is BV, then x_s is integrable.*

For each j the function $\gamma_j(x_s)$ is a.e. a derivative and is therefore measurable. Since $\{\gamma_j\}$ has property $N(\mathfrak{Y})$, x_s must be measurable by Theorem 2.1. Let I_0 be a cube containing R_0, and define $X'_{R'} = X_{R' \cdot R_0}$ for each figure $R' \subset I_0$. Then $X'_{R'}$ is BV if X_R is BV; and when x_s is extended to I_0 by setting $x'_s = x_s$ in R_0 and $x'_s = \theta$ elsewhere, then $X'_{R'}$ is $\mathfrak{Z}$-pseudo-differentiable to x'_s. Now let $\pi_m = [I_{m,1}, \cdots, I_{m,k_m}]$ $(m = 1, 2, \cdots)$ be a sequence of partitions of I_0 into non-degenerate subcubes $I_{m,j}$ such that $\lim_{m \to \infty} (\text{norm } \pi_m) = 0$. If s is in the interior of $I_{m,j}$, define x_s^m as $X'_{I_{m,j}} / | I_{m,j} |$; otherwise let $x_s^m = \theta$. For each m the function x_s^m is integrable, and since $X'_{R'}$ is $\{\gamma_j\}$-pseudo-differentiable to x'_s, we have a.e. in I_0

$$
(1) \quad \begin{aligned} \| x'_s \| &= \limsup_n | \gamma_n(x'_s) | = \limsup_n (\lim_m | \gamma_n(x_s^m) |) \\ &\leqq \limsup_n (\liminf_m \| \gamma_n \| \cdot \| x_s^m \|) \leqq \liminf_m \| x_s^m \|. \end{aligned}
$$

If $X'_{R'}$ is BV, then

$$
(2) \quad \int_{I_0} \| x_s^m \| \, ds \leqq \text{Var}(X'_{R'}; I_0) = \text{Var}(X_R; R_0) < \infty.
$$

From (1), (2), and Fatou's lemma it follows that $\| x'_s \|$ is majorized over I_0 by a summable function, $\liminf_m \| x_s^m \|$. But x'_s inherits measurability from x_s, so that $\| x'_s \|$ is measurable. Thus measurable x'_s has $\| x'_s \|$ summable, so that x'_s is integrable. The integrability of x_s is now obvious.

The proof of the preceding theorem is essentially a vindication of the following more general statement.

THEOREM 2.21. *Let x_s have a separable span $\mathfrak{Y}$. Then x_s is measurable if there exist a sequence $\{\gamma_j\}$ having property $N(\mathfrak{Y})$ and a sequence of functions $\{x_s^m\}$ such that*

(1) *$\gamma_j(x_s^m)$ is measurable for $j, m = 1, 2, \cdots$, and*

(2) *$\gamma_j(x_s) = \lim_m \gamma_j(x_s^m)$ a.e. for each j.*

If in addition

(3) *the indefinite integrals $\int_R \| x_s^m \| ds$ exist finitely or infinitely and*

$$\liminf_m \int_{R_0} \| x_s^m \| ds < \infty,$$

then x_s is integrable.

We are now ready to prove the theorem from which Theorem 2.5 will be easily derived. This is

THEOREM 2.3. *Suppose that X_R is an ABV function satisfying the following condition:*

(G) *Among the X_R-maximal sequences there is at least one, $\{\pi_n\}$, such that if $\mathfrak{W}$ is the separable X_R-span of $\{\pi_n\}$, then among the sequences having property $N(\mathfrak{W})$ there is at least one, $\{\delta_j\}$, such that X_R is $\{\delta_j\}$-pseudo-differentiable to the function θ_s.*

Then the real-valued ABV function $\Omega_R \equiv$ Var $(X_R\,;R)$ is singular.

Let $\pi_n = [R_{n,1}, \cdots, R_{n,k_n}]$ $(n = 1, 2, \cdots)$. Since $\{\delta_j\}$ has property $N(\mathfrak{W})$, where $\mathfrak{W}$ is the X_R-span of $\{\pi_n\}$, clearly for each element $X_{R_{n,i}}$ of $\mathfrak{W}$ there is a member $\delta_{n,i}$ of $\{\delta_j\}$ such that

$$\| X_{R_{n,i}} \| - \frac{1}{2^n k_n} \leqq | \delta_{n,i}(X_{R_{n,i}}) |. \tag{1}$$

Let $\Delta_R^n = \max [\, | \delta_{m,i}(X_R) |\,; 1 \leqq m \leqq n, 1 \leqq i \leqq k_m]$: then

$$0 \leqq \Delta_R^n \leqq \Delta_R^{n+1} \leqq \| X_R \|, \qquad R \subset R_0, \tag{2}$$

the last inequality on the right arising from the fact that $\| \delta_{m,i} \| \leqq 1$ for all i and m. From (2), each Δ_R^n is BV; moreover, Δ_R^n is convex, being the maximum of a finite number of convex functions $| \delta_{m,i}(X_R) |$. Since each Δ_R^n is BV and convex, the function $\Omega_R^n \equiv$ Var $(\Delta_R^n\,; R)$ is ABV; in addition, (2) implies that

$$0 \leqq \Omega_R^n \leqq \Omega_R^{n+1} \leqq \Omega_R, \qquad R \subset R_0, \tag{3}$$

so that the limit function $\Omega_R' \equiv \lim_n \Omega_R^n$ exists, is ABV, and satisfies the inequality

$$0 \leqq \Omega_R' \leqq \Omega_R, \qquad R \subset R_0. \tag{4}$$

From (1) on the other hand we can conclude that for any positive integer n

$$\Omega_{R_0}^n \geqq \sum_{i=1}^{k_n} \Delta_{R_{n,i}}^n \geqq \sum_{i=1}^{k_n} | \delta_{n,i}(X_{R_{n,i}}) | \geqq \sum_{i=1}^{k_n} \| X_{R_{n,i}} \| - 2^{-n};$$

if we take the limit as $n \to \infty$, this becomes, since $\{\pi_n\}$ is an X_R-maximal sequence,

$$\Omega'_{R_0} \equiv \lim_n \Omega^n_{R_0} \geqq \Omega_{R_0}. \tag{5}$$

For an arbitrary $R \subset R_0$ it now follows from (5) and (4) that

$$\Omega'_R = \Omega'_{R_0} - \Omega'_{R_0-R} \geqq \Omega_{R_0} - \Omega_{R_0-R} = \Omega_R,$$

and this combined with (4) results in the identity $\Omega'_R = \Omega_R$ for all $R \subset R_0$.

Thus $\{\Omega^n_R\}$ is a monotone increasing sequence of ABV functions converging to Ω_R. If it is shown that each Ω^n_R is differentiable a.e. to 0, that is, each Ω^n_R is singular, then Ω_R will necessarily be singular ([18], p. 94, Theorem 12.1, 3°) and the theorem will be established. But for each i and m the ABV function $\delta_{m,i}(X_R)$ is differentiable a.e. to 0, since by assumption X_R is $\{\delta_j\}$-pseudo-differentiable to $\theta_s \equiv \theta$. The set of singular functions being closed under the operations of addition and of taking the total variation ([18], p. 94, Theorem 12.1), it follows that the function

$$\Lambda^n_R \equiv \sum_{m=1}^{n} \sum_{i=1}^{k_m} \text{Var}\,(\delta_{m,i}(X_R); R)$$

must be singular for each n. Since

$$\Lambda^n_R \geqq \text{Var}\,(\sum_{m=1}^{n} \sum_{i=1}^{k_n} |\,\delta_{m,i}(X_R)|; R) \geqq \text{Var}\,(\Delta^n_R; R) = \Omega^n_R \geqq 0,$$

it is now seen that each Ω^n_R is also singular. This completes the demonstration.

THEOREM 2.4. *If X_R is ABV, then the following statements are equivalent*:
(2.41) *X_R satisfies condition* (G) *of Theorem* 2.3;
(2.42) *X_R is differentiable to θ a.e.*;
(2.43) *the function $\Omega_R \equiv$* Var $(X_R; R)$ *is singular*;
(2.44) *for each $\epsilon > 0$ there exists an open set E_ϵ such that $|E_\epsilon| < \epsilon$ and* Var $(X_R; E_\epsilon) =$ Var $(X_R; R_0)$.

Since Ω_R is ABV and non-negative, it is its own variation; the equivalence of (2.43) and (2.44) results from combining this fact with a well-known theorem in real-function theory ([18], p. 121, Theorem 7.8). That (2.42) follows from (2.43) is stated in III of §1; and (2.41) implies (2.43) by Theorem 2.3 above. Thus the only implication that remains to be established is that (2.42) implies (2.41).

If X_R is ABV, let $\{\pi_n\}$ be an X_R-maximal sequence and let $\mathfrak{W}$ be the separable X_R-span of $\{\pi_n\}$. According to a remark made in §1, there exists in $\bar{\mathfrak{X}}$ at least one sequence $\{\delta_j\}$ having property $N(\mathfrak{W})$. If X_R is differentiable to θ a.e., then clearly X_R is $\bar{\mathfrak{X}}$-differentiable to $\theta_s \equiv \theta$, and hence X_R is certainly $\{\delta_j\}$-pseudo-differentiable to θ_s. Thus X_R has property (G), and (2.42) implies (2.41).

The equivalence of (2.42) and (2.44) has been previously established in [10].

The next theorem, our principal result, is obtained almost immediately from Theorems 2.2 and 2.3.

THEOREM 2.5. *Let X_R be ABV. Suppose that x_s is a function having a separable span $\mathfrak{Y}$ and that X_R is $\{\gamma_j\}$-pseudo-differentiable to x_s, where $\{\gamma_j\}$ is one of the sequences in $\bar{\mathfrak{X}}$ having property $N(\mathfrak{Y})$. Then the integral $Y_R = \int_R x_s ds$ exists. In addition suppose, Z_R being the ABV function $X_R - Y_R$, that X_R is $\{\delta_j\}$-pseudo-differentiable to x_s for some sequence $\{\delta_j\}$ having property $N(\mathfrak{W})$, where $\mathfrak{W}$ is the separable Z_R-span of some Z_R-maximal sequence $\{\pi_n\}$. Then X_R is differentiable a.e. to x_s.*

The first assertion is merely a repeated statement of the previously established Theorem 2.2. The integral Y_R being differentiable a.e. to its integrand x_s, it follows that Z_R is the difference of two functions each of which is $\{\delta_j\}$-pseudo-differentiable to x_s. Since this implies that Z_R is $\{\delta_j\}$-pseudo-differentiable to θ_s, it is readily seen that Z_R satisfies condition (G) of Theorem 2.3 and therefore is differentiable to θ a.e. This proves the theorem, since $X_R = Y_R + Z_R$, where Y_R is differentiable a.e. to x_s.

As an immediate inference of Theorem 2.5 we have

THEOREM 2.6. *If X_R is ABV and x_s is separably-valued, and if there is a sequence $\{\gamma_j\}$ having property $N(\mathfrak{X})$ and such that X_R is $\{\gamma_j\}$-pseudo-differentiable to x_s, then x_s is integrable and X_R is differentiable a.e. to x_s.*

The next theorem is

THEOREM 2.7. *If X_R is ABV and if there exists a separably-valued function x_s such that X_R is $\{\gamma_j\}$-pseudo-differentiable to x_s for every sequence $\{\gamma_j\}$ in $\bar{\mathfrak{X}}$, then x_s is integrable and X_R is differentiable a.e. to x_s.*

Let $\mathfrak{Y}$ be the separable span of x_s. As we have noted in §1, the separability of $\mathfrak{Y}$ implies the existence of a sequence $\{\gamma_j\}$ in $\bar{\mathfrak{X}}$ that has property $N(\mathfrak{Y})$; since x_s is a $\{\gamma_j\}$-pseudo-derivative of X_R, x_s must be integrable. Now let $\{\pi_n\}$ be any Z_R-maximal sequence, where $Z_R = X_R - \int_R x_s ds$; the Z_R-span $\mathfrak{W}$ of $\{\pi_n\}$ is separable and hence there exists a sequence $\{\delta_j\}$ having property $N(\mathfrak{W})$. From the assumption in the theorem X_R must be $\{\delta_j\}$-pseudo-differentiable to x_s. All the hypotheses of Theorem 2.5 are now fulfilled, and the present conclusion follows.

Theorem 2.7 can be rephrased as

THEOREM 2.8. *If X_R is ABV and x_s is separably-valued, then if X_R is $\bar{\mathfrak{X}}$-pseudo-differentiable to x_s, it follows that x_s is integrable and is the derivative a.e. of X_R.*

This in turn permits the following improvement to be made on a result obtained previously ([15], p. 427).

THEOREM 2.9. *If X_R is ABV and $\mathfrak{X}$-differentiable to x_s, that is, X_R is weakly differentiable to x_s a.e., then x_s is integrable and X_R is differentiable a.e. to x_s.*[7]

Since x_s is an $\mathfrak{X}$-derivative, it must be equivalent to a separably-valued function y_s ([16], Theorem 1.2). The function y_s, being separably-valued and obviously an $\mathfrak{X}$-derivative of X_R, is a.e. the derivative of X_R by Theorem 2.8. Our conclusion now follows from the equivalence of y_s and x_s.

In concluding this section we remark that Theorem 2.9 immediately implies a result of Clarkson's [7], to the effect that if X_R is ABV and differentiable a.e., then the derivative function is integrable. The derivative is essentially separably-valued, and hence X_R is differentiable a.e. to a separably-valued integrable function x_s. It can easily be seen that the two functions X_R and x_s possess all the properties demanded in the hypotheses of Theorems 2.5, 2.7, 2.8, and 2.9. The conditions shown to be sufficient in those theorems are therefore also necessary in order that X_R be differentiable a.e.

3. **Weakly compact spaces.** In [12] and [15] it was shown that if R_0 is a linear figure and $\mathfrak{X}$ is a reflexive B-space, then the ABV function X_R is differentiable a.e. In the next theorem we remove from this statement the restriction that R_0 be linear and at the same time formally weaken the requirement of reflexiveness for $\mathfrak{X}$ by substituting weak compactness; that is, we suppose that in every bounded sequence in $\mathfrak{X}$ there is a subsequence converging weakly *to an element of* $\mathfrak{X}$. Accordingly we state

THEOREM 3.1. *If X_R is ABV and $\mathfrak{X}$ is weakly compact, then X_R is differentiable a.e. in R_0.*

Let I_0 be a cube containing R_0 and $\{\pi_n\}$ a sequence of partitions of I_0 into non-degenerate subcubes, with $\pi_n = [I_{n,1}, \cdots, I_{n,k_n}]$ and $\lim_n (\text{norm } \pi_n) = 0$. If s lies in the interior of $I_{n,i}$ and $R_0 \supset I_{n,i}$, define x_s^n to be $X_{I_{n,i}}/ \mid I_{n,i} \mid$; otherwise let $x_s^n = \theta$. From II of §1 and the weak compactness of $\mathfrak{X}$ it follows that for almost every $s \in R_0$ the sequence $\{x_s^n\}$ contains a subsequence converging weakly to an element x_s of $\mathfrak{X}$. The function x_s is then separably-valued since its span must ([1], p. 134) lie in the span of the sequence $\{x_s^n\}$ of simple functions. Moreover, from the definitions of $\{\pi_n\}$, $\{x_s^n\}$, and x_s it is seen that for each $\gamma \in \bar{\mathfrak{X}}$ the derivative of $\gamma(X_R)$ must coincide a.e. with $\gamma(x_s)$. Theorem 3.1 is now an immediate implication of Theorem 2.8.

Every reflexive space being necessarily weakly compact,[8] we have

THEOREM 3.2. *Any ABV function defined to a reflexive space is differentiable a.e.*

Since $\mathfrak{X}$ is reflexive if it is isomorphic to a uniformly convex space [14, 17], the following result of Clarkson's [7] is in turn a corollary of Theorem 3.2.

[7] This theorem furnishes another example of a weak property implying the corresponding strong property. For further examples see [9] and various general mean ergodic theorems that have lately appeared.

[8] A result obtained by Gantmakher and Šmulian [11].

THEOREM 3.3. *If X_R is ABV and has its values in a space $\mathfrak{X}$ that is isomorphic to a uniformly convex space, then X_R is differentiable a.e.*[9]

4. $\mathfrak{X}$ has a base. Suppose that $\mathfrak{X}$ has a base $\{x_i\}$, so that every $x \in \mathfrak{X}$ has a representation $x = \sum_{i=1}^{\infty} \zeta_i(x)x_i$, where $\{\zeta_i\}$ is a sequence in $\bar{\mathfrak{X}}$ that is independent of x. Let $\mathfrak{U}$ be the point-set product of the unit sphere of $\mathfrak{X}$ and the separable span $\mathfrak{Z}$ of $\{\zeta_i\}$. Each ABV function X_R defined to $\mathfrak{X}$ can be written as $X_R = \sum_{i=1}^{\infty} \zeta_i(X_R)x_i$, where for each i the function $\zeta_i(X_R)$ has a derivative a.e.; let $\dot{\zeta}_s^i$ be this derivative where it exists. If we consider a denumerable set dense in $\mathfrak{U}$ and consisting of finite linear combinations of the ζ_i's, the next theorem can be inferred directly from Theorem 2.6.

THEOREM 4.1. *Suppose the base $\{x_i\}$ has the property that*

$$\| x \| = \limsup_{\mathfrak{U}} | \gamma(x) |, \qquad x \in \mathfrak{X}.$$

Then if X_R is ABV and if $\dot{x}_s = \sum_{i=1}^{\infty} \dot{\zeta}_s^i x$ exists a.e. in R_0, X_R must be differentiable a.e. to $\dot{x}_s$.

From Theorem 4.1 the following result, obtained by Dunford and Morse [10], can be drawn.

THEOREM 4.2. *If $\mathfrak{X}$ has a base $\{x_i\}$ with the property*

(A) $$\limsup_n \left\| \sum_{i=1}^{n} a_i x_i \right\| < \infty \text{ implies that } \sum_{i=1}^{\infty} a_i x_i \text{ is convergent,}$$

then any ABV function X_R defined to $\mathfrak{X}$ is differentiable a.e.

In the proof given in [10] two preliminary facts are established:

(i) there is no loss of generality in assuming that $\{x_i\}$ has the property

(B) *the inequality* $\left\| \sum_{i=1}^{n} a_i x_i \right\| \leqq \left\| \sum_{i=1}^{n+1} a_i x_i \right\|$ *holds for any constants* $a_1, \cdots, a_{n+1}$;

(ii) properties (A) and (B) imply that $\dot{x}_s = \sum_{i=1}^{n} \dot{\zeta}_s x_i$ exists a.e.

In view of (ii) and Theorem 4.1, the present theorem will be justified if the inequality $\| x \| \leqq \limsup_{\mathfrak{U}} | \gamma(x) |$ is shown to hold for every x (the reverse inequality is obvious). To do this fix x, take $\gamma \in \bar{\mathfrak{X}}$ such that $\| \gamma \| = 1$ and $\| x \| = | \gamma(x) |$, and consider the element $\gamma_n = \sum_{i=1}^{n} \gamma(x_i)\zeta_i$ in $\mathfrak{Z}$. Property (B) shows that $\| \gamma_n \| \leqq 1$, so that γ_n must also be in $\mathfrak{U}$. The desired inequality now follows from this and from the fact that $\| x \| = | \gamma(x) | = \lim_n | \gamma_n(x) |$.

[9] For the case of a linear figure R_0 it is noted in [14] that this result of Clarkson's is a consequence of the theorem of Gelfand's cited at the beginning of this section.

5. **Separability assumptions and their application to the case of a general $\mathfrak{X}$.** Suppose that $\mathfrak{X} = [x]$ is the adjoint of a separable space $\mathfrak{Y} = [y]$, and let $\{y_j\}$ be a sequence dense in the unit sphere of $\mathfrak{Y}$. For each $y \in Y$ define the functional $\gamma_y(x) = x(y)$ over $\mathfrak{X}$; then $\mathfrak{Y}' = [\gamma_y]$ is a subset of $\bar{\mathfrak{X}}$ and forms a total set of linear functionals over $\mathfrak{X}$. If X_R and x_s are defined to $\mathfrak{X}$, we shall say that X_R is *$\mathfrak{Y}$-(pseudo-)differentiable to x_s* if X_R is $\mathfrak{Y}'$-(pseudo-)differentiable to x_s in the sense of our previous definitions. Writing γ_j for the element γ_{y_j} of $\mathfrak{Y}'$ and noting that $\{\gamma_j\}$ has property $N(\mathfrak{X})$ and is therefore total over $\mathfrak{X}$, the following amplification of Theorem 4 of [12] can now be stated.

THEOREM 5.1. *Let X_R be defined to such a space $\mathfrak{X}$. If X_R has its difference quotients bounded at almost every point in R_0, and if $\gamma_j(X_R)$ is differentiable a.e. in R_0 for each j, then X_R has a unique $\mathfrak{Y}$-derivative x_s, and x_s is measurable if it is essentially separably-valued.*

Hence if X_R is ABV and defined to the adjoint of a separable space $\mathfrak{Y}$, then X_R has a unique $\mathfrak{Y}$-derivative[10] *x_s; moreover, if x_s is essentially separably-valued, then x_s is integrable and X_R is differentiable a.e. to x_s.*

From the assumptions made in the theorem there exists a set S in the interior of R_0 having the three properties: (i) $| S | = | R_0 |$; (ii) at each point of S the difference quotients of X_R are bounded; and (iii) at each point of S the function $\gamma_j(X_R)$ is differentiable for every j. Let $\{x_n\}$ be a sequence of difference quotients of X_R defined by non-degenerate cubes closing down on a fixed point s in S. Then $\limsup_n \| x_n \| < \infty$, and $\lim_n \gamma_j(x_n) = \lim_n x_n(y_j)$ exists for every j where $\{y_j\}$ is dense in $\mathfrak{Y}$. This implies that $\{x_n\}$ converges to an x_s in $\mathfrak{X}$, in the sense that for every y we have

$$\gamma_y(x_s) = x_s(y) = \lim_n x_n(y) = \lim_n \gamma_y(x_n), \tag{1}$$

and in particular

$$\gamma_j(x_s) = \lim_n \gamma_j(x_n) = \frac{d}{ds}\gamma_j(X_R) \tag{2}$$

for every j by virtue of (iii). Thus if x_s' is the limit of any other sequence of difference quotients closing down on the point s, from (2) we have $\gamma_j(x_s) = \frac{d}{ds}\gamma_j(X_R) = \gamma_j(x_s')$ for every j, and hence $x_s = x_s'$ since $\{\gamma_j\}$ is total over $\mathfrak{X}$. This together with (1) implies that for every y the derivative of $\gamma_y(X_R)$ exists at s and has the value $\gamma_y(x_s)$. Since $| S | = | R_0 |$, it is now clear that X_R is $\mathfrak{Y}'$-differentiable to x_s; and since $\mathfrak{Y}'$ is total over $\mathfrak{X}$, this $\mathfrak{Y}'$-derivative must be unique.

[10] It is to be noted that "$\mathfrak{Y}$-derivative" can not be replaced here by the stronger term "$\mathfrak{X}$-derivative"; see (B) of §7 and Theorem 2.9. The "weak derivative" ascribed to X_R in Theorem 4 of [12] must be understood as the $\mathfrak{Y}$-derivative. The difference quotients of X_R converge weakly a.e. as functionals over $\mathfrak{X}$, but not necessarily do they converge weakly a.e. as elements of $\mathfrak{X}$.

If x_s is essentially separably-valued, then if we disregard values on a set of measure zero, it is separably-valued and yet still remains the $\mathfrak{Y}'$-derivative of X_E. Since the sequence $\{\gamma_j\}$ has property $N(\mathfrak{X})$, Theorem 2.2 now yields that x_s is measurable. Thus the first part of the theorem is justified.

The fact that X_E has a unique $\mathfrak{Y}$-derivative x_s if X_E is ABV follows from the preceding when it is recalled that an ABV function has its difference quotients bounded a.e. in R_0 and that $\gamma(X_E)$ is a.e. differentiable for each $\gamma \in \bar{\mathfrak{X}}$. The last statement in the theorem is a direct consequence of Theorem 2.6 since X_E is $\{\gamma_j\}$-pseudo-differentiable to x_s, where $\{\gamma_j\}$ has property $N(\mathfrak{X})$.

Because the values of x_s all lie in $\mathfrak{X}$, an obvious corollary ([12], Theorem 3) is

Corollary 5.11. *If the space $\mathfrak{X}$ of Theorem 5.1 is separable, then any ABV function X_E defined to $\mathfrak{X}$ is differentiable a.e. to its $\mathfrak{Y}$-derivative x_s.*

These results still have application even in the case in which X_E assumes values in an arbitrary space $\mathfrak{Z}$. In this connection we may point out

Theorem 5.2. *Any BV function Z_s defined from a linear interval to an arbitrary space $\mathfrak{Z}$ must have a separable span $\mathfrak{W}$.*

If R_0 is a linear interval $a \leqq s \leqq b$ and Z_s is a BV function defined from the points of R_0 to an arbitrary space $\mathfrak{Z}$, then if we utilize a certain parameter transformation $s = \sigma(t)$ ([10], §4), a function $X_t = Z_{\sigma(t)}$ can be obtained which satisfies a Lipschitz condition and whose functional values include those of Z_s. Since X_t is continuous over a linear interval, its span is separable, and hence Z_s has a separable span $\mathfrak{W}$ in $\mathfrak{Z}$.

Thus if Z_s is any BV function defined from a linear interval to an arbitrary space $\mathfrak{Z}$, there must be in $\bar{\mathfrak{Z}}$ a sequence $\{\gamma_j\}$ having property $N(\mathfrak{W})$. Letting $\mathfrak{Y}$ be the span of the sequence $\{\gamma_j\}$ in $\bar{\mathfrak{Z}}$, Z_s can be considered as defined to the adjoint $\bar{\mathfrak{Y}}$ of the separable space $\mathfrak{Y}$. As such, it must have a unique $\mathfrak{Y}$-derivative. If this $\mathfrak{Y}$-derivative is essentially separably-valued in $\bar{\mathfrak{Y}}$, then Z_s, considered as defined to $\bar{\mathfrak{Y}}$, must be differentiable a.e. Since $\mathfrak{Y}$ is the span of a sequence having property $M(\mathfrak{W})$, it can be shown further that Z_s is then differentiable a.e. when considered as a function defined to $\mathfrak{Z}$.

6. Remarks on the differentiation of abstract integrals.

In this section we should like to make a comment or two concerning the differentiation of certain abstract integrals that have been constructed for functions having the points in R_0 for their domain and an arbitrary B-space $\mathfrak{X}$ for their range. Only the following three distinct definitions of integrability, listed in decreasing order of generality, will be considered.

Definition 6.1. *x_s is $(\mathfrak{X})$ integrable* [9, 12, 16] *if $\gamma(x_s)$ is summable Lebesgue for every $\gamma \in \bar{\mathfrak{X}}$ and if for each measurable set $E \subset R_0$ there exists an element $x_E \in \mathfrak{X}$ such that $\gamma(x_E) = \int_E \gamma(x_s)\, ds$ for all $\gamma \in \bar{\mathfrak{X}}$. The integral $(\mathfrak{X}) \int_E x_s\, ds$ of x_s over E is by definition this element x_E.*

DEFINITION 6.2. *x_s is (D) integrable* [8, 16] *if it is measurable and ($\mathfrak{X}$) integrable. The (D) integral of x_s over E is taken to be* $(\mathfrak{X}) \int_E x_s\, ds$.

DEFINITION 6.3. *x_s is integrable if it is Bochner integrable.*

If x_s is integrable, then it is (D) integrable [8, 16], and if it is (D) integrable, then it is ($\mathfrak{X}$) integrable [16]; moreover, if x_s is integrable according to any two of these definitions, then the two integrals of x_s coincide over every measurable set [3, 8, 16]. All of these integrals are absolutely continuous and completely additive [3, 5, 8, 16]; in addition, every Bochner integral is BV, differentiable a.e., and also *absolutely additive*, that is, $\sum_{n=1}^{\infty} \left\| \int_{E_n} x_s\, ds \right\| < \infty$ whenever $\{E_n\}$ are disjoint measurable sets. The next theorem, the proof of which involves a simple application of Theorem 2.8, serves to make more precise the distinctions between the three integrals.

THEOREM 6.4. *If x_s is (D) integrable and $X_E \equiv (D) \int_E x_s\, ds$, then the following conditions are all equivalent:*

(6.5) $\qquad X_E$ *is absolutely additive;*

(6.6) $\qquad X_E$ *is BV;*

(6.7) $$\int_{E_0} \| x_s \|\, ds < \infty;$$

(6.8) $\qquad x_s$ *is integrable.*

Each of these conditions implies that

(6.9) $\qquad X_E$ *is differentiable a.e. to x_s.*

Since x_s is supposed measurable, the real-valued function $\| x_s \|$ must be measurable, so that the integral in (6.7) always exists, either finitely or infinitely. It is known [5] that (6.8) implies (6.5), (6.6), and (6.9), and that for a measurable x_s the conditions (6.7) and (6.8) are equivalent [5]. There remain only the proofs that (6.5) and (6.6) each implies (6.8).

Suppose that x_s is (D) integrable and that $X_E \equiv (D) \int_E x_s\, ds = (\mathfrak{X}) \int_E x_s\, ds$ is BV. Then $\gamma(X_E) = \int_E \gamma(x_s)\, ds$ holds for every γ, and $\gamma(X_E)$ is differentiable a.e. to $\gamma(x_s)$ since it is the Lebesgue integral of $\gamma(x_s)$. The ABV function X_E is then $\bar{\mathfrak{X}}$-pseudo-differentiable to x_s. Since x_s is measurable by assumption, it is essentially separably-valued. From Theorem 2.8 it follows that x_s is integrable. Thus (6.6) implies (6.8).

If x_s is (D) integrable, then [16] there exists a sequence $\{x_s^n\}$ of (D) integrable functions such that (i) each x_s^n has only a countable number of functional

values, and (ii) $\| x_s - x_s^n \| < n^{-1}$ holds uniformly a.e. in R_0. Assume that $\int_E x_s\, ds$ is absolutely additive; since for each n and each set E we have

$$\left\| \int_E x_s^n\, ds \right\| \leqq \left\| \int_E x_s\, ds \right\| + \frac{|E|}{n},$$

the integral $\int_E x_s^n\, ds$ must be absolutely additive for each n. It now follows that x_s^n is integrable. According to (i) there exist disjoint measurable sets E_{ni} $(i = 1, 2, \cdots)$ such that x_s^n has a constant value x_{ni} over E_{ni}, and $\sum_{i=1}^{\infty} E_{ni} = R_0$. Both x_s^n and $\| x_s^n \|$ are measurable, and since $\int_E x_s^n\, ds$ is absolutely additive, we can write

$$\int_{R_0} \| x_s^n \|\, ds = \sum_{i=1}^{\infty} \| x_{ni} \| \cdot | E_{ni} | = \sum_{i=1}^{\infty} \| x_{ni} \cdot | E_{ni} | \| = \sum_{i=1}^{\infty} \left\| \int_{E_{ni}} x_s^n\, ds \right\| < \infty.$$

Thus the measurable function x_s^n has $\| x_s^n \|$ summable, and x_s^n must be integrable. Combining this with (ii) we obtain the conclusion that x_s is a.e. the uniform limit of integrable functions, and therefore x_s is itself integrable.

The hypothesis of (D) integrability cannot be replaced by the weaker assumption of $(\mathfrak{X})$ integrability. Example 1 of [3] gives a non-measurable $(\mathfrak{X})$ integrable function whose $(\mathfrak{X})$ integral satisfies (6.5), (6.6), and (6.7), but not (6.8) or (6.9). Another example, the seventh, in [3] shows that (6.9) is not equivalent to any one of the conditions (6.5)–(6.8) even for (D) integrable functions.

7. **Two concluding remarks.** (A) In the second part of the proof of Lemma 1 of [10] there is implicitly established this proposition: *if X_R is a real-valued BV convex function*[11] *of figures, then X_R satisfies condition* (2.44) *if it satisfies condition* (2.42). This leads immediately to the following extension of a theorem which is well known for ABV real-valued functions.

THEOREM 7.1. *For real-valued BV convex functions of figures the conditions* (2.42), (2.43), *and* (2.44) *are all equivalent.*

From the above remark (2.42) implies (2.44). And since the ABV function $\Omega_R \equiv \mathrm{Var}\,(X_R\,;\,R)$ is non-negative and hence identical with its own variation function, it follows that (2.44) implies (2.43) since Theorem 7.1 is known to hold for ABV functions. The remaining implication that (2.42) results from (2.43) was proved in the introduction.

If we utilize Theorem 7.1, the argument that established Theorem 2.3 may be repeated to justify the following generalization.

[11] Such functions are closely allied to what Banach has called *normal* functions [2]. If X_R is real-valued convex and BV, then $-| X_R |$ is normal, and hence (loc. cit.) is differentiable a.e. On the other hand, every normal function is the difference of two non-negative convex BV functions.

THEOREM 7.2. *Let X_R be a convex BV function having its values in an arbitrary B-space, and suppose that X_R satisfies the following condition:*

(G′) *Among the X_R-maximal sequences there is at least one, $\{\pi_n\}$, such that if $\mathfrak{Y}$ is the separable X_R-span of $\{\pi_n\}$, then among the sequences having property $N(\mathfrak{Y})$ there is at least one, $\{\gamma_j\}$, such that $\gamma_j(X_R)$ is convex for each j and X_R is $\{\gamma_j\}$-pseudo-differentiable to θ_s.*

Then the real-valued ABV function $\Omega_R \equiv \mathrm{Var}\,(X_R\,;R)$ is singular.

A combination of Theorems 7.1 and 7.2 leads to

THEOREM 7.3. *If X_R is convex and BV, then* (2.42), (2.43), *and* (2.44) *are equivalent conditions, and X_R satisfies all three if it satisfies condition* (G′) *of Theorem* 7.2.

(B) Since the property of being essentially separably-valued is obviously a necessary condition that x_s be the derivative a.e. of an ABV function (see, for example, the proof of Theorem 3.1), it is reasonable to include this property in any set of sufficient conditions, as we have done in Theorems 2.5–2.8 and in Theorem 5.1 of §5. The question might arise, however, as to whether or not a set of sufficient conditions that includes this assumption still remains sufficient when the assumption is omitted. A single example answers this in the negative for all the above cited theorems. Let $\mathfrak{X}$ be the non-separable space consisting of the bounded sequences of real numbers, and consider the example given by Clarkson ([7], p. 414) of an additive function X_R of linear figures that is defined to $\mathfrak{X}$ and satisfies a Lipschitz condition, yet is nowhere differentiable. Here $\mathfrak{X}$ is the adjoint of the separable space $\mathfrak{Y}$ composed of absolutely convergent series, so that by Theorem 5.1 X_R has a unique $\mathfrak{Y}$-derivative x_s. On inspecting the actual example, we see that x_s fails to be essentially separably-valued and hence fails to be integrable; yet X_R and its $\mathfrak{Y}$-derivative x_s satisfy all the remaining hypotheses in Theorems 2.5–2.8 and the theorems of §5.

REFERENCES

1. S. BANACH, *Théorie des Opérations Linéaires*, Warsaw, 1932.
2. S. BANACH, *Sur une classe de fonctions d'ensemble*, Fund. Math., vol. 6(1924), pp. 170–180.
3. G. BIRKHOFF, *Integration of functions with values in a Banach space*, Trans. Amer. Math. Soc., vol. 38(1935), pp. 357–378.
4. S. BOCHNER, *Absolut-additiv abstrakte Mengenfunktionen*, Fund. Math., vol. 21(1933), pp. 211–213.
5. S. BOCHNER, *Integration von Funktionen, deren Werte die Elemente eines Vektorraumes sind*, Fund. Math., vol. 20(1933), pp. 262–276.
6. S. BOCHNER AND A. E. TAYLOR, *Linear functionals on certain spaces of abstractly-valued functions*, Annals of Math., vol. 39(1938), pp. 913–944.
7. J. A. CLARKSON, *Uniformly convex spaces*, Trans. Amer. Math. Soc., vol. 40(1936), pp. 396–414.
8. N. DUNFORD, *Integration of abstract functions*, Bull. Amer. Math. Soc., vol. 42(1936), p. 178 (abstract).
9. N. DUNFORD, *Uniformity in linear spaces*, Trans. Amer. Math. Soc., vol. 44(1938), pp. 305–353.

10. N. Dunford and A. P. Morse, *Remarks on the preceding paper of James A. Clarkson*, Trans. Amer. Math. Soc., vol. 40(1936), pp. 415–420.
11. V. Gantmakher and V. Šmulian, *Sur les espaces dont la sphère unitaire est faiblement compacte*, Comptes Rendus de l'Acad. des Sc. de l'URSS, vol. 17(1937), pp. 91–94.
12. I. Gelfand, *Zur Theorie abstrakter Funktionen*, Comptes Rendus de l'Acad. des Sc. de l'URSS, vol. 17(1937), pp. 243–245.
13. H. Hahn, *Über lineare Gleichungssysteme in linearen Räumen*, Journal für die r. u. a. Mathematik, vol. 157(1927), pp. 214–229.
14. D. Milman, *On some criteria for the regularity of spaces of type* (B), Comptes Rendus de l'Acad. des Sc. de l'URSS, vol. 20(1938), pp. 243–246.
15. B. J. Pettis, *A note on regular Banach spaces*, Bull. Amer. Math. Soc., vol. 44(1938), pp. 420–428.
16. B. J. Pettis, *On integration in vector spaces*, Trans. Amer. Math. Soc., vol. 44(1938), pp. 277–304.
17. B. J. Pettis, *A proof that every uniformly convex space is reflexive*, this Journal, vol. 5(1939), pp. 249–253.
18. S. Saks, *Theory of the Integral*, Warsaw-Lwów, 1937.

University of Virginia.

NON-COMMUTATIVE ARITHMETIC

By R. P. Dilworth

1. **Introduction and summary.** The problem of determining the conditions that must be imposed upon a system having a single associative and commutative operation in order to obtain unique factorization into irreducibles has been studied by A. H. Clifford [1],[1] König [1], and Ward [2]. The more general problem of determining similar conditions for the non-commutative case has been treated by M. Ward [1]. However, the conditions given by Ward are more stringent than those satisfied by actual instances of non-commutative arithmetic, for example, quotient lattices and non-commutative polynomial theory (Ore [1, 2]). Moreover, in both of these instances the factorization is unique only up to a similarity relation, and instead of a single operation of multiplication the additional operations G. C. D. and L. C. M. are involved.[2] Accordingly, we shall concern ourselves with the arithmetic of a non-commutative multiplication defined over a lattice.

As the decomposition of lattice quotients gives an important instance of non-commutative arithmetic, we shall summarize here a few of the fundamental ideas of Ore's theory (Ore [1]). Let Σ be the set of quotients[3]

$$\alpha = \frac{a_1}{a_2}, \qquad a_2 \supset a_1, \qquad a_1, a_2 \in L,$$

where L is a lattice in which the ascending chain condition holds. If $\beta = b_1/b_2$, we define $(\alpha, \beta) = (a_1, b_1)/(a_2, b_2)$, $[\alpha, \beta] = [a_1, b_1]/[a_2, b_2]$. With these definitions Σ is a lattice which is modular or distributive if and only if L is modular or distributive. Ore defines the product $\alpha \cdot \beta$ only for elements $\alpha, \beta \in \Sigma$ such that $a_2 = b_1$, in which case $\alpha \cdot \beta = a_1/b_2$. Let us set $\alpha \Leftrightarrow \beta$ if and only if $a_2 = b_1$, so that a necessary and sufficient condition for the existence of the product $\alpha \cdot \beta$ is that $\alpha \Leftrightarrow \beta$. Although the relation $\Leftrightarrow$ is neither reflexive nor symmetric, it is in a certain sense transitive since

$$\text{(1)} \qquad \text{if } \alpha \Leftrightarrow \beta \text{ and } \gamma \Leftrightarrow \delta, \text{ then } \gamma \Leftrightarrow \beta \text{ implies } \alpha \Leftrightarrow \delta.$$

Furthermore, the relation is preserved under union and cross-cut; that is,

$$\text{(2)} \qquad \alpha \Leftrightarrow \beta,\ \gamma \Leftrightarrow \delta \text{ implies } (\alpha, \gamma) \Leftrightarrow (\beta, \delta),\ [\alpha, \gamma] \Leftrightarrow [\beta, \delta].$$

Received September 2, 1938.

[1] The numbers in brackets refer to the references at the end of the paper.

[2] In this regard note that the necessary and sufficient conditions for unique factorization in the commutative case are stated in their most elegant form in terms of the G. C. D. operation (König [1]).

[3] Our inclusion is the reverse of Ore's.

Also

(3) $$\alpha \asymp \beta,\ \beta \asymp \gamma \text{ implies } \alpha \asymp \beta\cdot\gamma,\ \alpha\cdot\beta \asymp \gamma.$$

In the abstract theory given below we shall choose (1), (2), and (3) as the defining properties of the abstract relation $\asymp$.

When we have a commutative multiplication, the connection of the multiplication with the lattice operations automatically makes the lattice modular (in fact, distributive (Ward-Dilworth [1, 2])). If the multiplication is non-commutative, the lattice need not be modular; however, the assumption of the modular condition is essential since we shall prove that it is one of the necessary and sufficient conditions for arithmetic in a non-commutative semigroup. In particular, our results show the importance of the modularity[4] of a non-commutative polynomial domain in determining its arithmetical properties.

We conclude by stating our fundamental decomposition theorem for the elements of a lattice Σ with a multiplication having the properties of §2.

Decomposition Theorem. *Each element a of Σ not equal to a unit has a decomposition into irreducible elements. If there are two such decompositions*

$$a = p_r p_{r-1} \cdots p_2 p_1 = q_s q_{s-1} \cdots q_2 q_1,$$

then $r = s$ and the p's and q's are similar in pairs.

2. **The multiplication.**[5] Let Σ be a lattice in which the ascending chain condition holds and let i denote its unit element. Consider in Σ a relation $\asymp$ having the following properties:

T1. For each $a \in \Sigma$ there are elements $a', a'' \in \Sigma$ such that $a \asymp a'$, $a'' \asymp a$.

T2. $a \asymp b$, $a = c$, and[6] $b = d \rightarrow c \asymp d$.

T3. $a \asymp b$, $c \asymp d \rightarrow (a, c) \asymp (b, d)$, $[a, c] \asymp [b, d]$.

T4. If $a \asymp b$, $c \asymp d$, then $c \asymp b \rightarrow a \asymp d$.

Definition 2.1. Let L_a denote the set of elements x such that $x \asymp a$.

By T1 and T3, L_a is non-empty and closed with respect to union and cross-cut. Hence L_a is a sublattice of Σ. Thus with each element a of Σ we associate a sublattice L_a.

In a similar manner we may associate with each element a of Σ a sublattice S_a defined as the set of all x's such that $a \asymp x$.

Theorem 2.1. *The lattices L_a and L_b are either disjoint or they are identical.*

Proof. If L_a and L_b are not disjoint, they have an element c in common. Let x be an arbitrary element of L_a. Then $x \asymp a$, $c \asymp a$, $c \asymp b$, and hence $x \asymp b$ by T3. Similarly, each element of L_b belongs to L_a.

[4] That a non-commutative polynomial domain is modular can be easily seen from the fact that the degree of a polynomial is a rank function over the lattice in the sense of Birkhoff (Birkhoff [1], Ore [2]).

[5] See Ward-Dilworth [1, 2] for lattice notation.

[6] We shall use $\rightarrow$ to denote "implies".

Clearly, a similar result holds for S_a and S_b.

Let now L_s and $L_{s'}$ be the L-lattices corresponding to s, $s' \in S_a$. Then $a \in L_s$, $L_{s'}$ and hence L_s and $L_{s'}$ are identical by Theorem 1.1. *Thus to each S-lattice we can associate an L-lattice L, where L is the L-lattice corresponding to an arbitrary element of S; and conversely, to each element of L corresponds the S-lattice S.*

Definition 2.2. We write $a \sim b$ if a and b belong to the same L-lattice.

$\sim$ is clearly an equivalence relation in Σ with the L-lattices as equivalence classes.

We consider now a multiplication over Σ having the following properties:

M00. *To each pair of elements a, b such that $a \mathrel{\subset\!\supset} b$, there is ordered a unique element ab, the product of a and b.*

M0.[7] $a = b \rightarrow ac = bc, da = db$.

M1. *With each $a \in \Sigma$ there exists an element $u_a \mathrel{\subset\!\supset} a$ such that $u_a a = a$.*

M2. $a \supset ba$.

M3. $(a, b)c = (ac, bc)$.

M4. $(ab)c = a(bc)$.

M5. $ac = bc \rightarrow a = b$.

M6. $a \mathrel{\subset\!\supset} b, b \mathrel{\subset\!\supset} c \rightarrow a \mathrel{\subset\!\supset} bc, ab \mathrel{\subset\!\supset} c$.

From M00-M6 follow

$$\begin{aligned} &a \mathrel{\subset\!\supset} b, c \mathrel{\subset\!\supset} ab \rightarrow c \mathrel{\subset\!\supset} a; \\ &a \mathrel{\subset\!\supset} b, ab \mathrel{\subset\!\supset} c \rightarrow b \mathrel{\subset\!\supset} c. \end{aligned} \tag{2.1}$$

Proof. $u_a \mathrel{\subset\!\supset} a$ by M1. Hence $u_a \mathrel{\subset\!\supset} ab$ by M6. But $c \mathrel{\subset\!\supset} ab$ and thus $c \mathrel{\subset\!\supset} a$ by T4. A similar proof gives the second statement.

$$u_a \text{ is the unit element of } L_a. \tag{2.2}$$

Proof. Let $x \in L_a$. Then $a \supset xa \rightarrow u_a a = a = (a, xa) = (u_a a, xa) = (u_a, x)a$ by M1, M2, M3. Hence $u_a = (u_a, x)$ by M5.

$$a \supset b \rightarrow ac \supset bc \text{ if } a \mathrel{\subset\!\supset} c, b \mathrel{\subset\!\supset} c \text{ by M3.} \tag{2.3}$$

Let ${}_aL$ denote the L-lattice to which a belongs. Let ${}_au$ denote its unit element. Then we have

$$a\,{}_au = a. \tag{2.4}$$

Proof. We have $a \mathrel{\subset\!\supset} {}_au$ since if $a \mathrel{\subset\!\supset} x$, then ${}_au \mathrel{\subset\!\supset} x$ and ${}_aux = x$ by (2.2) and M1. But then $a \mathrel{\subset\!\supset} {}_aux$, and hence $a \mathrel{\subset\!\supset} {}_au$ by (2.1). Now $ax = a({}_aux) = (a\,{}_au)x \rightarrow a = a\,{}_au$ by M4 and M5.

Definition 2.3. b is said to *divide a on the right* if there is an element $x \in \Sigma$ such that $a = xb$.

The element x of Definition 2.3 is unique by M5 and is called the quotient of a and b.

[7] In this and the remaining postulates the statements are assumed to hold if and only if all the products appearing in the statements exist.

DEFINITION 2.4. If b divides $[a, b]$ on the right, we write $a \ominus b$ and denote the quotient of $[a, b]$ and b by $a \cdot b^{-1}$.

THEOREM 2.2. *The quotient $a \cdot b^{-1}$ has the following properties:*

R1. $$a \supset (a \cdot b^{-1})b;$$

R2. $$a \supset xb \rightarrow a \cdot b^{-1} \supset x.$$

Proof. R1 is clear from Definition 2.4. Let $a \supset xb$ so that $[a, b] \supset xb$ by M2. Then

$$(a \cdot b^{-1})b \supset xb \rightarrow (a \cdot b^{-1}, x)b = ((a \cdot b^{-1})b, xb) = (a \cdot b^{-1})b.$$

Hence $(a \cdot b^{-1}, x) = a \cdot b^{-1}$ by M5.

Since R1 and R2 are the defining properties of the residual (Ward-Dilworth [1, 2]), we shall call $a \cdot b^{-1}$ the residual of b with respect to a. It exists only if $a \ominus b$.

We note that $a \ominus a$ and $ba \ominus a$ since $[a, a] = u_a a$ and $[a, ba] = ba$. Hence the residuals $a \cdot a^{-1}$ and $ba \cdot a^{-1}$ always exist.

(2.5) $$a \cdot a^{-1} = u_a .$$

(2.6) $$(ab) \cdot b^{-1} = a.$$

(2.7) $$a \ominus c, b \ominus c \rightarrow [a, b] \ominus c.$$

(2.8) $$[a, b] \cdot c^{-1} = [a \cdot c^{-1}, b \cdot c^{-1}] \textit{ if } a \ominus c, b \ominus c.$$

(2.9) $$a \supset b \textit{ if and only if } a \cdot b^{-1} = u_b .$$

3. **The decomposition theory.** Throughout this section we make the following assumptions:

A1. $a \sim b \rightarrow a \ominus b$.[8]

A2. Σ *is modular.*

As consequences of A1 we have

(3.1) $$a \ominus c, b \ominus c \rightarrow a \cdot c^{-1} \ominus b \cdot c^{-1}.$$

Proof. Since $a \cdot c^{-1} \asymp c$, $b \cdot c^{-1} \asymp c$, we have $a \cdot c^{-1} \ominus b \cdot c^{-1}$ by A1.

(3.2) $$a \sim b \rightarrow a \ominus (a, b).$$

(3.3) $$(ba) \cdot c^{-1} = (b \cdot (c \cdot a^{-1})^{-1})(a \cdot c^{-1}) \textit{ if } b \asymp a, c \ominus a, a \ominus c, ba \ominus c.$$

[8] A1 is much stronger than necessary. However, since the weaker formulations are more complicated and artificial, and since the methods of proof remain essentially the same, we adopt the present formulation. If the proofs are examined, the various weaker conditions will be readily apparent to the reader (as, for example, the set given in §5). Also A1 is always satisfied in the important instances of the theory.

Proof. $((ba)\cdot c^{-1})c = [ba, c] = [ba, c, a] = ([ba, c]\cdot a^{-1})a = [b, c\cdot a^{-1}]a$ by (2.8). Now $ba\cdot a^{-1} \ominus c\cdot a^{-1}$ by (3.1) and hence $b \ominus c\cdot a^{-1}$ by (2.6). Hence

$$((ba)\cdot c^{-1})c = ((b\cdot(c\cdot a^{-1})^{-1})(c\cdot a^{-1}))a = (b\cdot(c\cdot a^{-1})^{-1})((c\cdot a^{-1})a)$$
$$= (b\cdot(c\cdot a^{-1})^{-1})[a, c] = (b\cdot(c\cdot a^{-1})^{-1})((a\cdot c^{-1})c) = ((b\cdot(c\cdot a^{-1})^{-1})(a\cdot c^{-1}))c$$

by M4 and Definition 2.4.

Definition 3.1. If $a' = a\cdot b^{-1}$, where $a \sim b$ and $(a, b) = {}_a u$, we say that a' is conjugate to a.

Definition 3.2. a is *similar* to b if there exists a chain of elements $a = a_0, a_1, \cdots, a_n = b$ such that either a_i is conjugate to a_{i+1} or a_{i+1} is conjugate to a_i (Ore [1]).

The relation of similarity is clearly reflexive, symmetric, and transitive.

Definition 3.3. An element $p \in \Sigma$ is irreducible if $p \neq {}_p u$, and if $x \supset p$, $x \sim p \rightarrow x = {}_p u$ or $x = p$.

(3.4) *If p is irreducible, then $p \not\supset a$ and $p \sim a \rightarrow (p, a) = {}_p u$.*

(3.5) *If p is irreducible and $p \supset ab$, $p \sim b$, $p \not\supset b$; then $p' \supset a$, where p' is conjugate to p.*

Proof. Take $p' = p\cdot b^{-1}$.

Theorem 3.1. *An element conjugate to an irreducible element is an irreducible element.*

Proof. Let p be an irreducible element, and let $p' = p\cdot a^{-1}$, where $p \sim a$ and $(p, a) = {}_p u$. Let $x \supset p'$ with $x \sim p'$. Then $x \supset p\cdot a^{-1}$ and $xa \supset (p\cdot a^{-1})a = [a, p]$ by (2.3), Definition 2.4. Hence $xa = (xa, [a, p]) = [a, (xa, p)]$ by A2. Thus $x = (xa)\cdot a^{-1} = [a, (xa, p)]\cdot a^{-1} = (xa, p)\cdot a^{-1}$ by (2.6), (2.8), (2.9). Now $a \Leftrightarrow d$ for some d by T1 and hence $p'a \Leftrightarrow d$, $xa \Leftrightarrow d$ by M6. Thus we have $xa \sim p'a = (p\cdot a^{-1})a = [a, p]$. But since $p \sim a$, $[a, p] \sim p$ and hence $xa \sim p$. $(xa, p) \supset p$ gives $(xa, p) = {}_p u$ or p since p is irreducible; and hence $x = {}_p u\cdot a^{-1} = u_a = {}_p u$ or $x = p\cdot a^{-1} = p'$. This proves the theorem.

Theorem 3.2. *If an irreducible p is conjugate to an element p', then p' is an irreducible.*

Proof. Let $p = p'\cdot a^{-1}$, where $p' \sim a$ and $(p', a) = {}_{p'} u$, and let $x \supset p'$, $x \sim p'$. Then $x\cdot a^{-1} \supset p'\cdot a^{-1}$ by (2.8), A1 and thus $x\cdot a^{-1} \supset p$. Also $x\cdot a^{-1} \sim p$ since $x\cdot a^{-1} \Leftrightarrow a$ and $p \Leftrightarrow a$. Hence we have either (i) $x\cdot a^{-1} = {}_p u$ or (ii) $x\cdot a^{-1} = p$. If (i) holds, then $x\cdot a^{-1} = u_a$ since $p \Leftrightarrow a$ and hence $x \supset a$ by (2.9). But $x \supset p'$ by hypothesis, hence $x \supset (a, p') = {}_{p'} u$. And since $x \sim p'$, $x = {}_{p'} u$. If (ii) holds, we have $(x\cdot a^{-1})a = (p'\cdot a^{-1})a$ or $[x, a] = [p', a]$ by Definition 2.4. But then $x \supset p' \supset [x, a]$ and by A2, $p' = [x, (p', a)] = [x, {}_{p'} u] = x$. Hence either $x = {}_{p'} u$ or $x = p'$, and thus p' is an irreducible. This completes the proof.

Theorem 3.3. *Every element similar to an irreducible element is irreducible.*

Proof. The theorem is clear from Theorems 3.1 and 3.2 and Definition 3.2.

THEOREM 3.4. *Let a' be conjugate to $a = a_k a_{k-1} \cdots a_2 a_1$, then $a' = a'_k a'_{k-1} \cdots a'_2 a'_1$, where a'_i is conjugate to a_i.*

Proof. Suppose the theorem is true for every product of $k - 1$ elements and let $a' = a \cdot b^{-1}$, $a \sim b$, $(a, b) = {}_b u$. Then $a \ominus b$, $a_1 \ominus b$, $b \ominus a_1$ since $a \sim b$ and $a_1 \sim b$. Thus $a' = ((a_k \cdots a_2) \cdot (b \cdot a_1^{-1})^{-1})(a_1 \cdot b^{-1})$ by (3.3). Now $a_1 \sim b$ and $(a_1, b) \supset (a, b) = {}_b u$ which gives $(a_1, b) = {}_b u$. Hence $a'_1 = a_1 \cdot b^{-1}$ is conjugate to a_1. Let $b' = b \cdot a_1^{-1}$ and $s = a_k \cdots a_2$. Then $b' \sim s$ since $b' \simeq a_1$ and $s \simeq a_1$. Now $(s, b')a_1 = (sa_1, b'a_1) = (a, (b \cdot a_1^{-1})a) = (a, [b, a_1]) = [a_1, (b, a)] = [a_1, {}_a u] = a_1$ by Definition 2.4 and A2. Hence $(s, b') = a_1 \cdot a_1^{-1} = {}_{b'} u$. Thus $s \cdot b'^{-1}$ is conjugate to s and hence $s \cdot b'^{-1} = a'_k a'_{k-1} \cdots a'_2$ by hypothesis. Substitution gives $a' = a'_k \cdots a'_1$. The theorem is therefore proved.

We now prove the fundamental

UNIQUENESS THEOREM. *If an element $a \in \Sigma$ has two representations as a product of irreducibles*

$$a = p_r p_{r-1} \cdots p_2 p_1 = q_s q_{s-1} \cdots q_2 q_1,$$

then $r = s$ and the p's and q's are similar in pairs.

Proof.[9] Let

$$a = p_r p_{r-1} \cdots p_2 p_1 = q_s q_{s-1} \cdots q_2 q_1. \tag{1}$$

If $p_1 = q_1$, this factor may be canceled. If $p_1 \neq q_1$, let k be the first number such that $q_1 \supset p_k p_{k-1} \cdots p_2 p_1$; then $q_1 \not\supset p_{k-1} \cdots p_1$ and $p_{k-1} \cdots p_1 \sim q_1$. Hence $(q_1, p_{k-1} \cdots p_1) = {}_{q_1} u$ by (3.4). But then $q_1 \cdot (p_{k-1} \cdots p_1)^{-1} \supset p_k$ and $q'_1 = q_1 \cdot (p_{k-1} \cdots p_1)^{-1}$ is conjugate to q_1 and thus is an irreducible by Theorem 3.2. Hence $q'_1 = p_k$. Now $p_k p_{k-1} \cdots p_1 = (q_1 \cdot (p_{k-1} \cdots p_1)^{-1}) p_{k-1} \cdots p_1 = [q_1, p_{k-1} \cdots p_1] = ((p_{k-1} \cdots p_1) \cdot q_1^{-1}) q_1$ by Definition 2.4. Hence $p_k p_{k-1} \cdots p_1 = p'_{k-1} \cdots p'_1 q_1$ by Theorem 3.4 and p'_i is conjugate to p_i $(i = 1, \cdots, k-1)$. Substituting this result in (1) and canceling q_1, we may treat the resulting expression in the same manner. Thus we find $r = s$ and the p's and q's similar in pairs.

Concerning the existence of a decomposition into irreducibles we have the

EXISTENCE THEOREM. *If the descending chain condition holds for the right factors of $a \neq {}_a u$, then a has a decomposition into irreducible elements.*

Proof. If a is not an irreducible, then there is an element $a_1 \neq {}_a u$ such that $a_1 \supset a$, $a_1 \sim a$ and $a_1 \neq a$. But then $a = (a \cdot a_1^{-1}) a_1$ by A1. If a_1 is not an irreducible, we have an $a_2 \neq {}_a u$ such that $a_2 \supset a_1$, $a_2 \sim a_1$, and $a_2 \neq a_1$. Then $a = (a \cdot a_1^{-1})(a \cdot a_2^{-1}) a_2$. Thus we get a chain of elements $a \subset a_1 \subset a_2 \subset \cdots$ which must break off giving an irreducible element p_1 such that $a = bp_1$. But if b is not an irreducible, $b = b_1 p_2$. Since $p_1 \supset p_2 p_1 \supset \cdots$ is a descending chain of factors of a, it must break off giving a decomposition $a = p_k p_{k-1} \cdots p_2 p_1$. The proof of the theorem is complete.

[9] This proof is essentially that given by Ore [2] for non-commutative polynomials.

We note that the descending chain condition for the factors of an element of Σ does not follow from the ascending chain condition in Σ, as in the commutative case. However, it does follow from the ascending chain condition in Σ', where Σ' is the lattice of left union and cross-cut if they exist.

As examples of the abstract theory let Σ be the lattice N of a non-commutative polynomial domain. Then for every a and b, $a \Leftrightarrow b$ and $a \ominus b$ so that T1-T4, M6 and A1 are trivially satisfied. The relations of similarity and conjugacy are identical. Furthermore, in N the irreducible elements are those elements whose only right divisors are the elements themselves and the elements of the fundamental field.

More generally the above results apply to any non-commutative domain of integrity having a Euclidean algorithm.

Again if we interpret Σ to be the quotient lattice Q of a lattice L, we have $A \Leftrightarrow B$, where $A = a_1/a_2$, $B = b_1/b_2$ if and only if $a_2 = b_1$ in which case $AB = a_1/b_2$. Furthermore $u_A = a_1/a_1$, ${}_Au = a_2/a_2$. Postulates T1-T4 are clearly satisfied by the relation $\Leftrightarrow$, and it is readily verified that the multiplication satisfies M00-M6. We have $A \ominus B$ if and only if $a_2 \supset b_2$ in which case $A \cdot B^{-1} = [a_1, b_1]/b_1$. We observe that $A \sim B$ if and only if $a_2 = b_2$ so that A_1 is satisfied. If we start with a modular lattice L, then Σ is modular and A2 is satisfied. The irreducible elements are those quotients p for which p_2 covers p_1.[10]

4. **The arithmetic of a semigroup.** Let S be a semigroup of elements $a, b, c, \cdots$ and unit element i such that each pair of elements a, b has a G. C. D. (a, b). Then if the ascending chain condition holds in S, a and b have an L. C. M. defined as the G. C. D. of those elements which both a and b divide. As in §2 we define $a \cdot b^{-1} = [a, b]/b$.

Definition 4.1. If $a' = a \cdot b^{-1}$, where $(a, b) = i$, we say that a' is conjugate to a.

Definition 4.2. a is *similar* to b if there exists a chain of elements $a = a_0, \cdots, a_n = b$ such that either a_i is conjugate to a_{i+1} or a_{i+1} is conjugate to a_i.

We have then the following fundamental theorem:

Theorem 4.1. *Let S be a semigroup with G. C. D. and L. C. M. operations. (S is thus a lattice with respect to G. C. D. and L. C. M.) Then the following*

[10] As another example, consider the set M of all finite matrices for which the number of rows is greater than or equal to the number of columns over a non-commutative ring R with unit element. A subset A of M is called an ideal if the matrices of A (i) all have the same number of rows and the same number of columns, (ii) are closed under addition, and (iii) are closed under multiplication by all square matrices for which the product exists. We write $A \Leftrightarrow B$ if the matrices of A have the same number of columns as the matrices of B have rows. The product of A and B is defined only if $A \Leftrightarrow B$ and is the ideal generated by the products of the matrices of A with those of B. With a suitable definition of union and cross-cut the set Σ of ideals of M satisfies T1-T4, M00-M4, M6, A2. Moreover, if we give a similar definition of left ideals in M and R is a non-commutative domain of integrity for which every left ideal is principal, then the set Σ of left ideals of M satisfies T1-A2 and is an instance of our abstract theory. A detailed account of these systems will be given in another paper.

three conditions are necessary and sufficient that each element not equal to i of S be expressible as a product of irreducibles unique up to similarity:[11]

(i) *the ascending chain condition in S;*

(ii) *the descending chain condition for the right factors of each element in S;*

(iii) *the modular condition in S.*

Proof. The sufficiency of conditions (i)-(iii) follows from the results of §3. For since the product of any two elements always exists, T1-T4 are trivially satisfied. M00-M6 are readily verified and A1 is trivially true since $a \ominus b$ for every a and b. (iii) gives A2. Hence the existence and uniqueness theorems of §3 hold.

Suppose now that each element not equal to i of S is uniquely (up to similarity) expressible as a product of irreducibles. We define $\rho(i) = 0$, $\rho(a) = s$ if $a = p_s p_{s-1} \cdots p_2 p_1$. Then $\rho(a) = 0$ if and only if $a = i$ and $\rho(a) = 1$ if and only if a is an irreducible. Furthermore, $\rho(ab) = \rho(a) + \rho(b)$ since if $a = p_{\rho(a)} \cdots p_1$ and $b = q_{\rho(b)} \cdots q_1$, then $ab = p_{\rho(a)} \cdots p_1 q_{\rho(b)} \cdots q_1$. Hence $a \supset b$ and $a \neq b$ implies that $\rho(a) < \rho(b)$. It follows that the ascending chain condition holds in S and the descending chain condition holds for the factors of each element in S.

We note that if a' is similar to a, $\rho(a') = \rho(a)$.

Let a and b be any two elements of S. We have then $a = a_1(a, b)$, $b = b_1(a,b)$, where $(a_1, b_1) = i$. Then

$$[a, b] = [a_1(a, b), b_1(a, b)] = [a_1, b_1](a, b) = (a_1 \cdot b_1^{-1}) b_1(a, b) = a_1' b_1(a, b),$$

where a_1' is similar to a_1. Then

$$\rho([a, b]) = \rho(a_1') + \rho(b_1) + \rho((a, b)) = \rho(a_1) + \rho(b_1) + \rho(a, b).$$

But $\rho(a) = \rho(a_1) + \rho((a, b))$, so that $\rho(a_1) = \rho(a) - \rho((a, b))$. Similarly, $\rho(b_1) = \rho(b) - \rho((a,b))$. Hence $\rho([a,b]) = \rho(a) + \rho(b) - \rho((a,b))$ or $\rho([a,b]) + \rho((a,b)) = \rho(a) + \rho(b)$. Thus ρ is a rank function over S in the sense of Birkhoff (Birkhoff [1], p. 447) and S is modular by Birkhoff's result. Hence conditions (i)-(iii) are satisfied.

5. **Properties of the L-lattices.** Using the notations of §3, we make

DEFINITION 5.1. The unit elements of the L-lattices are called the units of Σ.

Let now $a_1, a_2 \in L$, $a_1', a_2' \in L'$, where L and L' are any two L-lattices. Then if $a_1, a_2 \simeq x_1$, $a_1', a_2' \simeq x_2$, we have $(a_1, a_1') \simeq (x_1, x_2)$ and $(a_2, a_2') \simeq (x_1, x_2)$. Hence (a_1, a_1') and (a_2, a_2') belong to the same L-lattice. We call this L-lattice to which all the unions of elements from L_1 and L_2 respectively belong the union of L_1 and L_2 and write (L_1, L_2). In a similar manner we define the cross-cut $[L_1, L_2]$ of two L-lattices. Hence we make the L-lattices into a lattice Σ_l. Σ_l will be modular if L is modular.

[11] By "up to similarity" we mean that the irreducibles appearing in the decompositions of similar elements are similar in pairs.

In general, the union of the unit elements of L_1 and L_2 will not be the unit element of (L_1, L_2). However, we prove

THEOREM 5.1. *If the descending chain condition holds in Σ, then the units of Σ are closed under union and cross-cut and form a lattice isomorphic to Σ_l.*

Proof. We prove first a necessary lemma.

LEMMA 5.1. *If the descending chain condition holds in Σ, then the only elements of Σ such that $x \Leftrightarrow x$ are the units of Σ.*

Proof of lemma. We note that $u \Leftrightarrow u$ for every unit u, since if $u \Leftrightarrow x$, then $x = ux$; and hence $u \Leftrightarrow u$ by M6. Now let $a \Leftrightarrow a$. Then the chain $a, a^2, a^3, \cdots$ must break off by the descending chain condition so that $a^{m+n} = a^n$ or $a^m = u_a$ by M5. But since $a \Leftrightarrow a$, $u_a = {}_au$ and hence $a^m = {}_au$. We have then $a^{m-1} = {}_au \cdot a^{-1} = {}_au$, and finally $a = {}_au$.

We continue with the proof of the theorem. Let u and u' be two units, so that $u \Leftrightarrow u$, $u' \Leftrightarrow u'$. Then $(u, u') \Leftrightarrow (u, u')$ and $[u, u'] \Leftrightarrow [u, u']$. Whence (u, u') and $[u, u']$ are units by Lemma 5.1. This completes the proof of the theorem.

The L-lattices have a number of interesting interrelations. We mention, however, only one:

THEOREM 5.2. *Let L be an arbitrary L-lattice and let $l \in L$. Then L has a sublattice isomorphic to L_l with l as the unit element.*

Proof. Let $x \in L_l$ and set up the correspondence $x \leftrightarrow xl$, where xl is clearly in L. Then $(x, y) \leftrightarrow (x, y)l = (xl, yl)$ and $[x, y] \leftrightarrow [x, y]l = [xl, yl]$. Furthermore, the correspondence is 1-1 by M5. Hence the theorem follows.

We next characterize the irreducibles of Σ in terms of the lattice properties of the L-lattices.

THEOREM 5.3. *The irreducibles of Σ are the divisor-free elements of the L-lattices.*

Proof. Let p be a divisor-free element of an L-lattice, and let $x \supset p$, $x \sim p$; then clearly $x = {}_pu$ or $x = p$. Conversely, if p is an irreducible, let p' be the divisor-free element of ${}_pL$ dividing p. Then $p' \supset p$, $p' \sim p$ and $p' \neq {}_pu$, and hence $p' = p$.

We note that Theorem 5.3 may not hold if we weaken postulate A1. For example, let us replace A1 by

B1. $$a \ominus b, \quad a \ominus c, \quad b \ominus c \rightarrow a \cdot c^{-1} \ominus b \cdot c^{-1}.$$

B2. $$a \supset b \quad \text{and} \quad b \ominus c \rightarrow a \ominus c.$$

B3. $$a \ominus b \quad \text{and} \quad a \sim b \rightarrow a \ominus (a, b).$$

We define conjugate elements and irreducible elements by

DEFINITION 5.1. If $a' = a \cdot b^{-1}$, where $a \sim b$, $a \ominus b$, $b \ominus a$, and $(a, b) = {}_au$, we say that a' is conjugate to a.

DEFINITION 5.2. p is an irreducible if $p \neq {}_pu$ and if $x \supset p, x \sim p, p \ominus x \rightarrow x = p$ or $x = {}_pu$.

With these definitions the proofs of the existence and uniqueness theorems follow, with some modifications as in §3. However, there may be irreducibles which are not divisor-free elements since we may have $x \supset p, x \neq p, {}_pu; x \sim p$, but $x \ominus p$ is not true.

We conclude this section with the investigation of the special case where each L-lattice and its corresponding S-lattice are identical. Then $\Leftrightarrow$ is an equivalence relation and the equivalence classes are multiplicatively closed sublattices. Let $a = a_s \cdots a_1 = b_r \cdots b_1$ be two decompositions of a. Then $a \Leftrightarrow a_1$, and $a \Leftrightarrow b_1$. But $a_s \cdots a_2 \Leftrightarrow a_1$ and $a_s \cdots a_3 \Leftrightarrow a_2$. Hence $a \Leftrightarrow a_s \Leftrightarrow \cdots \Leftrightarrow a_1 \Leftrightarrow b_r \Leftrightarrow \cdots \Leftrightarrow b_1$.

Thus this case reduces to that of Σ closed under multiplication.

6. **The commutative case.** In this section we investigate the consequences of assuming that the multiplication is commutative. Explicitly we assume

A3. $$a \Leftrightarrow b \rightarrow b \Leftrightarrow a, ab = ba.$$

We have then

(6.1) $$a \Leftrightarrow b, b \Leftrightarrow c \rightarrow a \Leftrightarrow c.$$

Proof. Since $a \Leftrightarrow b$, ab exists and $ab \Leftrightarrow c$. But $ab = ba \Leftrightarrow c$, whence $a \Leftrightarrow c$.

(6.2) $$a \Leftrightarrow a.$$

Hence $\Leftrightarrow$ is an equivalence relation giving equivalence classes $\{a\}, \{b\}, \cdots$. The L-lattices and the equivalence classes coincide. Furthermore, we note that each equivalence class is closed with respect to union, cross-cut, multiplication and residuation. We note also that $u_a = {}_au$.

THEOREM 6.1. *a' is conjugate to a if and only if $a' = a$.*

Proof. We prove first a series of lemmas.

LEMMA 6.1. *If $a \sim b \sim c$ and $b \supset c$, then $a \cdot c^{-1} \supset a \cdot b^{-1}$.*

Proof. The residuals exist by A1. Furthermore, $a \supset (a \cdot b^{-1})b \supset (a \cdot b^{-1})c \rightarrow a \cdot c^{-1} \supset a \cdot b^{-1}$ by R1 and R2.

LEMMA 6.2. $a \sim b \sim c \rightarrow a \cdot (b, c)^{-1} = [a \cdot b^{-1}, a \cdot c^{-1}]$.

Proof. $[a \cdot b^{-1}, a \cdot c^{-1}] \supset a \cdot (b, c)^{-1}$ by Lemma 6.1. But $a \supset ((a \cdot b^{-1})b, (a \cdot c^{-1})c) \supset [a \cdot b^{-1}, a \cdot c^{-1}](b, c)$. Hence $a \cdot (b, c)^{-1} \supset [a \cdot b^{-1}, a \cdot c^{-1}]$ and thus $a \cdot (b, c)^{-1} = [a \cdot b^{-1}, a \cdot c^{-1}]$.

LEMMA 6.3. $a \cdot {}_au^{-1} = a \cdot u_a^{-1} = a$.

Proof. $a \supset au_a \rightarrow a \cdot u_a^{-1} \supset a$. But $a = au_a \supset (a \cdot u_a^{-1})a \rightarrow a \supset a \cdot u_a^{-1}$. Hence $a = a \cdot u_a^{-1}$. Since $u_a = {}_au$, the lemma follows.

We continue with the proof of the theorem. Let $a' = a\cdot b^{-1}$, $a \sim b$, $(a, b) = {}_a u$. Then

$$a = a\cdot {}_a u^{-1} = a\cdot(a, b)^{-1} = [a\cdot a^{-1}, a\cdot b^{-1}] = [{}_a u, a\cdot b^{-1}] = a\cdot b^{-1} = a'$$

by Lemmas 6.2 and 6.3. The proof of the theorem is complete.

Now obviously any irreducible factor of an element belongs to the same equivalence class as the element itself. Furthermore, since multiplication is commutative, the ascending chain condition implies the descending chain condition for the factors of an element $a \in \Sigma$. Hence by the uniqueness and existence theorems and Theorem 6.1 each element not a unit in Σ is uniquely expressible as a product of irreducibles, the irreducibles belonging to the same equivalence class. Thus considered as a lattice, each equivalence class is a direct product of chain lattices; i.e., an arithmetical lattice (Ward [3]).

References

Garrett Birkhoff.

1. *On combination of subalgebras*, Proc. Cambridge Phil. Soc., vol. 29(1933), pp. 441–464.

A. H. Clifford.

1. *Arithmetic and ideal theory of abstract multiplication*, Bull. Am. Math. Soc., vol. 40(1934), pp. 326–330.

J. König.

1. *Algebraischen Grossen*, Leipzig, 1903, Chapter I.

O. Ore.

1. *Abstract algebra*, I, II, Annals of Math., vol. 36(1935), pp. 406–437; vol. 37(1936), pp. 265–292.
2. *Theory of non-commutative polynomials*, Annals of Math., vol. 34(1933), pp. 480–508.

M. Ward.

1. *Postulates for an abstract arithmetic*, Proc. Natl. Acad. Sci., vol. 14(1928), pp. 907–911.
2. *Conditions for factorization in a set closed under a single operation*, Annals of Math., vol. 36(1933), pp. 36–39.
3. *Structure residuation*, Annals of Math., vol. 39 (1938), pp. 558–568.

M. Ward and R. P. Dilworth.

1. *Residuated lattices*, Proc. Natl. Acad. Sci., vol. 24(1938), pp. 162–165.
2. *Residuated lattices*, to appear in the Trans. of the Amer. Math. Soc.

J. H. M. Wedderburn.

1. *Non-commutative domains of integrity*, Journal für Math., vol. 167(1931), pp. 129–141.

California Institute of Technology.

ASYMPTOTIC FORMS FOR A GENERAL CLASS OF HYPERGEOMETRIC FUNCTIONS WITH APPLICATIONS TO THE GENERALIZED LEGENDRE FUNCTIONS

By George E. Albert

1. **Introduction.** The classical differential equation of Jacobi [5][1]

$$(1) \quad (1 - z^2)y'' + \{\beta - \alpha - (\alpha + \beta + 2)z\}y' + \nu(\nu + \alpha + \beta + 1)y = 0$$

is solved by the pair of hypergeometric functions (to be designated as the Jacobi functions)

$$(2) \quad \begin{cases} Y_{\nu,1}^{(\alpha,\beta)}(z) = F(\nu + \alpha + \beta + 1, -\nu; \alpha + 1; \tfrac{1}{2}(1 - z)), \\ Y_{\nu,2}^{(\alpha,\beta)}(z) = [\tfrac{1}{2}(z - 1)]^{-\nu-\alpha-\beta-1} \\ \qquad\qquad \cdot F(\nu + \alpha + \beta + 1, \nu + \beta + 1; 2\nu + \alpha + \beta + 2; 2/(1 - z)). \end{cases}$$

In the following pages forms will be derived for the Jacobi functions (2) which are asymptotic with respect to the large parameter ν.

The Legendre functions of complex degree, order, and argument are defined in terms of the Jacobi functions (2) by the formulas

$$(3) \quad \begin{cases} P_\nu^\mu(z) = \dfrac{1}{\Gamma(1 - \mu)} \left(\dfrac{z + 1}{z - 1}\right)^{\frac{1}{2}\mu} Y_{\nu,1}^{(-\mu,\mu)}(z), \\ 2Q_\nu^\mu(z) = e^{\mu\pi i} \dfrac{\Gamma(\nu + 1)\Gamma(\nu + \mu + 1)}{\Gamma(2\nu + 2)} \left(\dfrac{z + 1}{z - 1}\right)^{\frac{1}{2}\mu} Y_{\nu,2}^{(-\mu,\mu)}(z); \end{cases}$$

see Hobson [3] or [4]. In virtue of these relations between the two classes of functions, asymptotic forms will be at hand for the Legendre functions for values of $|\nu|$ which are large in comparison with $|\mu|$, and conversely.

I. The Jacobi functions

2. **The normalization of the differential equation.** In the differential equation (1) the numbers ν, α, and β will be subject to the blanket restrictions that $|\nu|$ be large and $|\alpha|$, $|\beta|$ be bounded; otherwise they are general complex numbers. The variable z will be allowed to range over the unbounded complex plane, cut along the axis of reals from the point $z = 1$ to $z = -\infty$, with the exception of an arbitrarily small neighborhood of the point $z = -1$. The domain of z thus defined will be consistently designated by R_z. In virtue of the known continuation formulas for hypergeometric functions, the omission of any small neighborhood of the point $z = -1$ involves no loss of generality in the results to be obtained.

Received October 8, 1938; in revised form, March 22, 1939.

[1] Numbers in brackets refer to the bibliography at the end of the paper.

The transformations

$$z = 1 + \tfrac{1}{4}s^2, \tag{4}$$

$$y(s) = 2^{\alpha+\beta+2}[s^{2\alpha+1}(8 + s^2)^{\beta+1}]^{-\frac{1}{2}}u(s) \tag{5}$$

change the differential equation (1) into the form

$$u''(s) + \left\{\rho^2\phi^2(s) + \frac{\frac{1}{4} - \alpha^2}{s^2} + \chi(\alpha, \beta, s)\right\}u(s) = 0 \tag{6}$$

wherein

$$\begin{cases} \rho^2 = -[\nu + \frac{1}{2}(\alpha + \beta + 1)]^2, \\ \phi^2(s) = \frac{1}{2}(1 + \frac{1}{8}s^2)^{-1}, \\ \chi(\alpha, \beta, s) = \dfrac{\alpha^2 + \beta^2 - 1 + \frac{1}{8}\alpha^2 s^2}{8(1 + \frac{1}{8}s^2)^2}. \end{cases} \tag{7}$$

Under the transformation (4) a one-to-one correspondence is established between R_z and a domain R_s consisting in that half-plane of the variable s given by the inequalities

$$-\tfrac{1}{2}\pi < \arg s \leqq \tfrac{1}{2}\pi,$$

with the exception of certain small neighborhoods of the points $s = \pm i\sqrt{8}$.

For the sake of definiteness the following agreements will be made: the root in (5) and the root $\phi(s)$ will be chosen as those which have positive real values when s is positive real; the symbol α will denote that root of α^2 for which $-\frac{1}{2}\pi < \arg \alpha \leqq \frac{1}{2}\pi$; the parameter ρ will, for convenience, be chosen so that $-\frac{1}{2}\pi - \delta < \arg \rho \leqq \frac{3}{2}\pi - \delta$, $0 < \delta < \frac{1}{2}\pi$.[2]

The coefficients $\phi^2(s)$ and $\chi(\alpha, \beta, s)$ of the differential equation (6) are analytic, non-vanishing functions of s in the domain R_s, and $\chi(\alpha, \beta, s)$ is uniformly bounded with respect to the parameters α and β.

The solutions of a differential equation of the type (6) have been derived by Langer [6] for large values of $|\rho|$ and for values of the independent variable in a domain of the type R_s. It follows that the equation (1) admits of known solutions for large values of $|\nu|$ when z remains within the domain R_z.

It is convenient to introduce the complex quantities

$$\begin{cases} \Phi = 2^{-\frac{1}{2}}\displaystyle\int_0^s (1 + \tfrac{1}{8}s^2)^{-\frac{1}{2}}\,ds = \log[z + (z^2 - 1)^{\frac{1}{2}}], \\ \xi = \rho\Phi, \end{cases} \tag{8}$$

in which the logarithm has its principal value. The domain R_s is mapped biuniquely upon the domain R_Φ consisting of the half-strip of the Φ-plane specified by the relations

$$-\pi < I(\Phi) \leqq \pi, \qquad -\tfrac{1}{2}\pi < \arg \Phi \leqq \tfrac{1}{2}\pi,$$

[2] Henceforth the symbol δ will always denote such a number; usually it will be taken small.

from which some small neighborhoods of the points $\Phi = \pm i$ have been removed (it will be assumed that these neighborhoods are of a shape to make R_Φ convex). The function $\Phi(z)$ is analytic over R_z and non-vanishing except at the point $z = 1$.

In the sequel use will be made of the solutions derived in [6] for the type equation (6). It must therefore be shown that a hypothesis imposed by Langer upon the equation is fulfilled in the present instance. It will be required that[3]

$$\int \left| \frac{\theta(s)\, ds}{\phi(s)} \right| < M$$

be satisfied uniformly by some constant M along any straight path in R_Φ upon which $|s| \geqq N_1 > 0$. For the case in hand it is found that

$$\frac{\theta(s)}{\phi(s)} = (\tfrac{1}{4} - \alpha^2)\left\{\frac{1}{s^2(1 + \frac{1}{8}s^2)^{\frac{1}{2}}} - \frac{\phi}{\Phi^2}\right\} + \frac{\beta^2 - \alpha^2}{8(1 + \frac{1}{8}s^2)^{\frac{3}{2}}}.$$

It follows that

$$\frac{\theta(s)\, ds}{\phi(s)} = O\left(\frac{ds}{s^3}\right) + O\left(\frac{d\Phi}{\Phi}\right),$$

and the fulfillment of the condition is immediate. In fact, it is satisfied uniformly for all values of α and β in any preassigned bounded domain of those quantities.

3. **Asymptotic forms for the Jacobi functions.** In [6] separate forms are given for the solutions of (6) according as the variable z is inside or outside a circle of radius $c/|\nu|$ (c denotes some positive constant) about the point $z = 1$. This division of the values of z will be designated briefly by the relations $|\xi| \leqq N$ and $|\xi| > N$, respectively.

The function $Y_{\nu,1}^{(\alpha,\beta)}(z)$ given in (2) is easily identified in terms of a solution $y_1(z)$ of the equation (1) derived in [6], Theorem 1. The function $y_1(z)$ is uniquely characterized as that solution of the differential equation which, when $R(\alpha) > 0$, approaches a constant as z approaches unity, to a higher order than any other solution. This property is evidently possessed by the Jacobi function $Y_{\nu,1}^{(\alpha,\beta)}(z)$. It follows that $Y_{\nu,1}^{(\alpha,\beta)}(z)$ and $y_1(z)$ differ only by a constant factor. Simple computations involving the relation

$$\lim_{z \to 1} (z - 1)^{-\frac{1}{2}}\xi = \rho$$

and the form given in [6] for the solution $y_1(z)$ when $|\xi| \leqq N$ lead to the formula

$$(9) \qquad Y_{\nu,1}^{(\alpha,\beta)}(z) = 2^{\frac{1}{2}(\alpha+\beta-\frac{1}{2})}\Gamma(\alpha + 1)\{i[\nu + \tfrac{1}{2}(\alpha + \beta + 1)]\}^{-(\alpha+\frac{1}{2})} y_1(z).$$

The second Jacobi function $Y_{\nu,2}^{(\alpha,\beta)}(z)$ given in (2) is uniquely characterized as that solution of (1) for which

$$(10) \qquad \lim_{|z| \to \infty} z^{\nu+\alpha+\beta+1} Y_{\nu,2}^{(\alpha,\beta)}(z) = 2^{\nu+\alpha+\beta+1}.$$

[3] See [6], pp. 400, 405 for the definition of the quantity $\theta(s)$.

When $|z| \to \infty$, the variable ξ is in the region $|\xi| > N$. For such values of ξ the solutions given in [6] change as that variable crosses the boundaries of regions $\Xi^{(h)}$ specified by the inequalities

$$\Xi^{(h)}: \quad (h-1)\pi + \delta \leqq \arg \xi \leqq (h-1)\pi - \delta. \tag{11}$$

By the equations (8) it is seen that

$$\arg \xi = \tfrac{1}{2}\pi + \arg [\nu + \tfrac{1}{2}(\alpha + \beta + 1)] + \arg \Phi \tag{12}$$

and, moreover, when $|z| \to \infty$, $\arg \Phi \to 0$. It is thus evident that $\xi|_{|z|\to\infty}$ is in the region $\Xi^{(1)}$ when $-\frac{1}{2}\pi + \delta \leqq \arg [\nu + \frac{1}{2}(\alpha + \beta + 1)] \leqq \pi - \delta$, and in the region $\Xi^{(0)}$ when $-\pi - \delta < \arg [\nu + \frac{1}{2}(\alpha + \beta + 1)] \leqq \frac{1}{2}\pi - \delta$. In [6], Theorem 2, fundamental pairs of solutions $y_{h,j}(z)$ $(j = 1, 2)$, for any integer h, are deduced. For values of ξ in the region $\Xi^{(h)}$ these solutions are uniquely characterized by the relations

$$\begin{cases} \lim\limits_{|z|\to\infty} z^{\nu+\alpha+\beta+1} y_{h,1}(z) = 2^{-\nu-\frac{1}{2}(\alpha+\beta+\frac{1}{2})}\{1 + O(\nu^{-1})\}, \\ \lim\limits_{|z|\to\infty} z^{-\nu} y_{h,2}(z) = 2^{\nu+\frac{1}{2}(\alpha+\beta+\frac{1}{2})}\{1 + O(\nu^{-1})\}. \end{cases}$$

It is immediate that for large enough values of $|\nu|$, the Jacobi function is, to within a constant factor, the solution $y_{h,1}(z)$ with $h = 0$ or 1 according as $\xi|_{|z|\to\infty}$ is in $\Xi^{(0)}$ or $\Xi^{(1)}$, respectively. However, $y_{0,1}(z) \equiv y_{1,1}(z)$ by definition ([6], (11)). It follows that

$$Y_{\nu,2}^{(\alpha,\beta)}(z) = 2^{2\nu+\frac{1}{2}(\alpha+\beta)+\frac{1}{2}} y_{0,1}(z)\{1 + O(\nu^{-1})\}. \tag{13}$$

The substitution of the general forms given in [6] for the solutions $y_1(z)$ and $y_{0,1}(z)$ into (9) and (13) respectively yields asymptotic forms for the two functions under consideration which are valid

(A) (i) *in the entire complex plane of z from which some small neighborhood of the point $z = 1$ has been excluded and which has been cut by the relations* $-\pi < \arg (z \pm 1) \leqq \pi$;
(ii) *for values of ν limited by the cut* $-\pi - \delta < \arg [\nu + \frac{1}{2}(\alpha + \beta + 1)] \leqq \pi - \delta$;
(iii) *uniformly for all α and β in any preassigned bounded domain.*

The facts are incorporated in the theorem which follows.

THEOREM I. *Under the conditions* (A), *the Jacobi functions $Y_{\nu,1}^{(\alpha,\beta)}(z)$ and $Y_{\nu,2}^{(\alpha,\beta)}(z)$ admit of the representations, asymptotic as to ν, given by the formulas*

$$\begin{cases} Y_{\nu,1}^{(\alpha,\beta)}(z) = 2^{\frac{1}{2}(\alpha+\beta)}\Gamma(\alpha + 1)\{i[\nu + \frac{1}{2}(\alpha + \beta + 1)]\}^{-\alpha} \\ \qquad\qquad \cdot (z-1)^{-\frac{1}{2}(\alpha+\frac{1}{2})}(z+1)^{-\frac{1}{2}(\beta+\frac{1}{2})}\Phi^{\frac{1}{2}}\{J_\alpha(\xi) + R_1\}, \\ Y_{\nu,2}^{(\alpha,\beta)}(z) = \pi^{\frac{1}{2}} 2^{2\nu+\frac{1}{2}(\alpha+\beta)+1} e^{\frac{1}{2}(\alpha+1)\pi i}[\nu + \frac{1}{2}(\alpha + \beta + 1)]^{\frac{1}{2}} \\ \qquad\qquad \cdot (z-1)^{-\frac{1}{2}(\alpha+\frac{1}{2})}(z+1)^{-\frac{1}{2}(\beta+\frac{1}{2})}\Phi^{\frac{1}{2}}\{H_\alpha^{(1)}(\xi) + R_2\}, \end{cases} \tag{14}$$

in which the correction terms R_i $(i = 1, 2)$ have the forms

$$R_1 = \begin{cases} \xi^{\alpha+2} O(\nu^{-2}) & \textit{for} \quad |z - 1| \leqq c/|\nu|, \\ \Phi^{-\frac{1}{2}}\{e^{i\xi} O(\nu^{-1}) + e^{-i\xi} O(\nu^{-1})\} & \textit{for} \quad |z - 1| > c/|\nu|, \end{cases}$$

and

$$R_2 = \begin{cases} \xi^{-\alpha} O(\nu^{-1}) & \textit{if} \quad \alpha \neq 0, \quad \textit{for} \quad |z - 1| \leqq c/|\nu|, \\ (\log \xi) O(\nu^{-1}) & \textit{if} \quad \alpha = 0, \quad \textit{for} \quad |z - 1| \leqq c/|\nu|, \\ R_1 & \qquad\qquad\quad \textit{for} \quad |z - 1| > c/|\nu|. \end{cases}$$

The symbols $J_\alpha(\xi)$ and $H_\alpha^{(1)}(\xi)$ denote the Bessel functions of the first and third kinds respectively. The symbol c denotes some fixed positive constant; the variables ξ and Φ are as given explicitly in (8).

For values of z such that $|z - 1| > c/|\nu|$ the formulas (14) may be replaced by the alternative forms given in terms of elementary functions, [6]. These forms depend upon the region $\Xi^{(h)}$ of the variable ξ. Recalling (12) and the restrictions upon ρ and Φ, we see that the total range of arg ξ is given by $-\pi - \delta < \arg \xi \leqq \pi - \delta$. The regions $\Xi^{(h)}$ with indices $h = -1, 0, 1$ cover this range entirely. It is found that *when $|z - 1| > c/|\nu|$ the Jacobi functions admit of the forms*

$$(15) \quad \begin{cases} Y_{\nu,1}^{(\alpha,\beta)}(z) = \pi^{-\frac{1}{2}} 2^{\frac{1}{2}(\alpha+\beta-1)} \Gamma(\alpha + 1)[\nu + \frac{1}{2}(\alpha + \beta + 1)]^{-(\alpha+\frac{1}{2})}(z - 1)^{-\frac{1}{2}(\alpha+\frac{1}{2})} \\ \qquad\qquad\qquad \cdot (z + 1)^{-\frac{1}{2}(\beta+\frac{1}{2})}\{a_{1,1}^{(h)} e^{i\xi} + a_{1,2}^{(h)} e^{-i\xi}\}, \\ \tilde{Y}_{\nu,2}^{(\alpha,\beta)}(z) = 2^{2\nu+\frac{1}{2}(\alpha+\beta+1)}(z - 1)^{-\frac{1}{2}(\alpha+\frac{1}{2})}(z + 1)^{-\frac{1}{2}(\beta+\frac{1}{2})}\{a_{2,1}^{(h)} e^{i\xi} + a_{2,2}^{(h)} e^{-i\xi}\}, \end{cases}$$

with coefficients $a_{i,j}^{(h)}$ dependent upon the region $\Xi^{(h)}$ of the variable ξ as shown in Table I.[4]

TABLE I

h	$a_{1,1}^{(h)}$	$a_{1,2}^{(h)}$	$a_{2,1}^{(h)}$	$a_{2,2}^{(h)}$
1	$[e^{(\alpha+\frac{1}{2})\pi i}]$	[1]	[1]	0
0	$[e^{-(\alpha+\frac{1}{2})\pi i}]$	[1]	[1]	0
−1	$[e^{-(\alpha+\frac{1}{2})\pi i}]$	$[-e^{-2\alpha\pi i}]$	[1]	$[-2i \cdot \cos \alpha\pi]$

The representations (14) and (15) above are *analytic* functions of the variable z.

The formulas (15) with the coefficients $a_{i,j}^{(h)}$ as stated in Table I for the regions $\Xi^{(0)}$ and $\Xi^{(1)}$ of the variable ξ were given by Watson [8]. He applied the well known method of steepest descents to two hypergeometric functions

[4] Henceforth the symbol $[E]$ will always denote a function of the form $[E] = E + O(\nu^{-1}) + O(\xi^{-1})$.

which are equivalent to the functions (2). Simple algebraic manipulations suffice to reduce his formulas to the above indicated cases of (15). His statements of the regions of validity for the results are the same as those to be found in the special cases (b) and (c) of §4 here.

4. **The regions of validity.** The asymptotic forms (15) change as the variable ξ crosses the boundaries of the regions $\Xi^{(h)}$. It is to be noted that any two regions for consecutive values of h overlap in the major part of an upper or lower half-plane. In this common domain the apparently different forms are asymptotically equivalent. It follows that the actual line across which any two forms interchange may be placed arbitrarily within the common half-plane of their validity.

While the relations (11) describe the regions $\Xi^{(h)}$ in all their generality, it is desirable to rephrase their content in terms of the variations of z or ν when one of these quantities is fixed.

When $\arg \Phi$ is fixed, the boundaries of the regions $\Xi^{(h)}$ (as seen in the plane of ν) are radial lines issuing from the point $\nu = -\frac{1}{2}(\alpha + \beta + 1)$ which divide the ν-plane into sectors. Of particular importance are the following special cases:

(a) When $|z|$ is large, the entire right half-plane of ν is contained within $\Xi^{(0)}$ and $\Xi^{(1)}$ simultaneously while the upper and the lower half-planes are contained within $\Xi^{(1)}$ and $\Xi^{(0)}$, respectively.

(b) For any value of z such that $I(z) \geqq 0$, the admissible values of ν are distributed between $\Xi^{(1)}$ and $\Xi^{(0)}$ as follows:[5]

$$\Xi^{(1)}: \qquad -\tfrac{1}{2}\pi + \delta \leqq \arg \nu \leqq \pi - \delta,$$

$$\Xi^{(0)}: \qquad -\pi - \delta < \arg \nu \leqq -\delta.$$

(c) For any value of z such that $I(z) \leqq 0$ the important range $|\arg \{\nu + \frac{1}{2}(\alpha + \beta + 1)\}| \leqq \pi - \delta$ is portioned between $\Xi^{(1)}$ and $\Xi^{(0)}$ by the relations[5]

$$\Xi^{(1)}: \qquad \delta \leqq \arg \nu \leqq \pi - \delta,$$

$$\Xi^{(0)}: \qquad -\pi + \delta \leqq \arg \nu \leqq \tfrac{1}{2}\pi - \delta.$$

When $\arg \nu$ is fixed, the boundaries of the regions $\Xi^{(h)}$ as seen in R_z are the curvilinear arcs represented by $\arg \Phi =$ constant. In general these curves will leave the point $z = 1$ at the angle $2 \arg \Phi$ with the positive axis of reals and bend round in such a way as to cut the negative axis of reals in a point whose modulus is greater than unity. The arcs resemble logarithmic spirals. In general the entire domain R_z will be contained within two of the regions $\Xi^{(h)}$ $(h = -1, 0, 1)$. The following cases are of especial importance.

(d) When ν is a positive real number, the entire domain R_z with the exception of a small convex region enclosing the real interval $-1 \leqq z \leqq 1$ is con-

[5] Since $|\nu|$ is very large relative to $|\alpha|$ and $|\beta|$, one has $\arg (\nu + \frac{1}{2}(\alpha + \beta + 1)) \sim \arg \nu$.

tained within both $\Xi^{(0)}$ and $\Xi^{(1)}$. The upper and lower half-planes are contained within $\Xi^{(1)}$ and $\Xi^{(0)}$, respectively.

(e) When $0 \leqq \arg \nu \leqq \pi - \delta$, ξ may be taken in $\Xi^{(1)}$, and when $-\pi + \delta \leqq \arg \nu \leqq 0$, ξ may be taken in $\Xi^{(0)}$.

5. **The Jacobi polynomials.** The classical Jacobi polynomials are defined by the formula

$$P_n^{(\alpha,\beta)}(z) = \frac{\Gamma(n+\alpha+1)}{\Gamma(n+1)\Gamma(\alpha+1)} F(n+\alpha+\beta+1, -n; \alpha+1; \tfrac{1}{2}(1-z))$$

in which n is a positive integer [5]. Making use of the asymptotic formula for the gamma function and the note (d) of the previous section, one has

COROLLARY I. *Asymptotic forms for the Jacobi polynomials are obtained from the first formulas in* (14) *and* (15) *and the relation*

$$P_n^{(\alpha,\beta)}(z) \sim \frac{n^\alpha}{\Gamma(\alpha+1)} Y_{n,1}^{(\alpha,\beta)}(z).$$

When (15) *is used, the variable ξ is confined to the regions* $\Xi^{(h)}$ $(h = 0, 1)$.

Imposing the restrictions α and β real and positive, $-1 < z < 1$ with $z = \cos\theta$ upon the forms obtained from the corollary, one obtains the classical result of Darboux [2] and the more recent result of Szegö [7].

II. The generalized Legendre functions

6. **Asymptotic forms for $P_\nu^\mu(z)$ and $Q_\nu^\mu(z)$ when $|\nu|$ is large.** The formulas (3) constitute the most general definitions for the Legendre functions. For the special case in which z is confined to the real interval $-1 \leqq z \leqq 1$ it is customary to set $z = \cos\theta$ $(0 \leqq \theta \leqq \pi)$ and to redefine the functions by means of the formulas

$$(16)\quad \begin{cases} P_\nu^\mu(\cos\theta) = e^{\frac{1}{2}\mu\pi i} P_\nu^\mu(\cos\theta + 0\cdot i), \\ 2e^{\mu\pi i} Q_\nu^\mu(\cos\theta) = e^{-\frac{1}{2}\mu\pi i} Q_\nu^\mu(\cos\theta + 0\cdot i) + e^{\frac{1}{2}\mu\pi i} Q_\nu^\mu(\cos\theta - 0\cdot i). \end{cases}$$

For general values of z it is known that

$$(17)\quad \begin{cases} P_\nu^\mu(-z) = e^{\mp\nu\pi i} P_\nu^\mu(z) - \dfrac{2}{\pi}\sin(\nu+\mu)\pi\, e^{-\mu\pi i} Q_\nu^\mu(z), \\ Q_\nu^\mu(-z) = -e^{\pm\nu\pi i} Q_\nu^\mu(z), \end{cases}$$

in which the upper or the lower signs are to be taken according as $I(z) \gtrless 0$. When $z = \cos\theta$, these relations are replaced by

$$(17')\quad \begin{cases} P_\nu^\mu(-\cos\theta) = P_\nu^\mu(\cos\theta)\cos(\nu+\mu)\pi - \dfrac{2}{\pi} Q_\nu^\mu(\cos\theta)\sin(\nu+\mu)\pi, \\ Q_\nu^\mu(-\cos\theta) = -Q_\nu^\mu(\cos\theta)\cos(\nu+\mu)\pi + \dfrac{\pi}{2} P_\nu^\mu(\cos\theta)\sin(\nu+\mu)\pi. \end{cases}$$

These four identities allow the limitation of the variable z to its right half-plane. However, the restriction will not be imposed except where greater simplicity results.

From the formulas (3) and (14) one obtains the representations

$$
(18)\qquad \begin{cases} P_\nu^\mu(z) = e^{\frac{1}{2}\mu\pi i}\,\dfrac{(\nu+\frac{1}{2})^\mu\Phi^{\frac{1}{2}}}{(z^2-1)^{\frac{1}{2}}}\,\{J_{-\mu}\{i(\nu+\tfrac{1}{2})\Phi\}+R_1\}, \\[2ex] Q_\nu^\mu(z) = e^{\frac{1}{2}(\mu+1)\pi i}\,\pi^{\frac{1}{2}}\,2^{2\nu}\,\dfrac{\Gamma(\nu+1)\Gamma(\nu+\mu+1)}{\Gamma(2\nu+2)}\,\dfrac{(\nu+\frac{1}{2})^{\frac{1}{2}}\Phi^{\frac{1}{2}}}{(z^2-1)^{\frac{1}{2}}} \\ \qquad\qquad\qquad\qquad\qquad\qquad\cdot\{H_{-\mu}^{(1)}\{i(\nu+\tfrac{1}{2})\Phi\}+R_2\}, \end{cases}
$$

in which the correction terms R_i $(i = 1, 2)$ are as given in Theorem I with the substitution $\alpha = -\mu$ and the variables ξ and Φ are explicitly

$$
\xi = i(\nu+\tfrac{1}{2})\Phi, \qquad \Phi = \log\{z + (z^2-1)^{\frac{1}{2}}\}.
$$

Similarly, when $|\xi| > N$, the formulas (3) and (15) lead to the representations

$$
(19)\qquad \begin{cases} P_\nu^\mu(z) = (2\pi)^{-\frac{1}{2}}(\nu+\tfrac{1}{2})^{\mu-\frac{1}{2}}(z^2-1)^{-\frac{1}{4}}\{b_{1,1}^{(h)}e^{i\xi}+b_{1,2}^{(h)}e^{-i\xi}\}, \\[2ex] Q_\nu^\mu(z) = 2^{2\nu+1}e^{\mu\pi i}\,\dfrac{\Gamma(\nu+1)\Gamma(\nu+\mu+1)}{\Gamma(2\nu+2)(z^2-1)^{\frac{1}{4}}}\,\{b_{2,1}^{(h)}e^{i\xi}+b_{2,2}^{(h)}e^{-i\xi}\}, \end{cases}
$$

in which the coefficients $b_{i,j}^{(h)}$ are dependent upon the region $\Xi^{(h)}$ of the variable ξ and are obtained from the $a_{i,j}^{(h)}$ given in Table I by the substitution $\alpha = -\mu$.

THEOREM II. *The representations, asymptotic as to* ν, *for the generalized Legendre functions* $P_\nu^\mu(z)$ *and* $Q_\nu^\mu(z)$, *valid when* $-\pi < \arg(z \pm 1) \leqq \pi$ *and* $-\pi - \delta < \arg \nu \leqq \pi - \delta$, *are given in general by the formulas* (18). *For values of* z *such that* $|z - 1| > c/|\nu|$, *the forms* (19) *hold. The forms* (18) *and* (19) *are valid uniformly for all values of* μ *in any bounded domain.*

Important specializations of the forms (18) are obtained upon applying the definitions (16). The notation $(\cos\theta \pm 0\cdot i)$ of (16) implies the limiting values: $\Phi = \pm i\theta$, $(\cos^2\theta - 1)^{-\frac{1}{4}} = e^{\mp\frac{1}{4}\pi i}(\sin\theta)^{-\frac{1}{2}}$, in which the upper or the lower signs are to be taken together. Easy calculations, if the identities $J_{-\mu}(se^{\pi i}) = e^{-\mu\pi i}J_{-\mu}(s)$, $2iY_{-\mu}(s) = H_{-\mu}^{(1)}(s) - H_{-\mu}^{(2)}(s)$ between Bessel functions are employed, lead to the representations

$$
(20)\qquad \begin{cases} P_\nu^\mu(\cos\theta) = (\nu+\tfrac{1}{2})^\mu\left(\dfrac{\theta}{\sin\theta}\right)^{\frac{1}{2}}J_{-\mu}\{(\nu+\tfrac{1}{2})\theta\} + R_1', \\[2ex] Q_\nu^\mu(\cos\theta) = -\dfrac{\pi}{2}(\nu+\tfrac{1}{2})^\mu\left(\dfrac{\theta}{\sin\theta}\right)^{\frac{1}{2}}Y_{-\mu}\{(\nu+\tfrac{1}{2})\theta\} + R_2', \end{cases}
$$

in which the correction terms R_i' $(i = 1, 2)$, have the forms

$$R_1' = \begin{cases} \theta^{2-\mu}O(1) & \text{for } 0 < \theta \leqq c/|\nu|, \\ e^{\nu\theta}O(\nu^{\mu-\frac{3}{2}}) + e^{-\nu\theta}O(\nu^{\mu-\frac{3}{2}}) & \text{for } \theta > c/|\nu| \text{ and for general complex values of } \nu, \\ O(\nu^{\mu-\frac{3}{2}}) & \text{for } \theta > c/|\nu| \text{ and real values of } \nu; \end{cases}$$

$$R_2' = \begin{cases} \theta\cdot O(\nu^{\mu-1}) & \text{if } \mu \neq 0 \text{ and } 0 < \theta \leqq c/|\nu|, \\ (\log \nu\theta)O(\nu^{-1}) & \text{if } \mu = 0 \text{ and } 0 < \theta \leqq c/|\nu|, \\ R_1' & \text{otherwise.} \end{cases}$$

The forms (20) *are valid uniformly in μ for values in any bounded domain and for θ in the indicated portions of the range $0 < \theta \leqq \pi - \delta$.*

A similar treatment of the formulas (19) leads to the classical asymptotic formulas for the functions under consideration.

7. **Asymptotic forms for $P_\nu^\mu(z)$ and $Q_\nu^\mu(z)$ when $|\mu|$ is large.** Utilizing the transformation

$$(z^2 - 1)(t^2 - 1) = 1,$$

Whipple [9] has obtained relations between the Legendre functions of degree ν and order μ and those of degree $-\mu - \frac{1}{2}$ and order $-\nu - \frac{1}{2}$ which hold when $R(z) \geqq 0$. Written in forms convenient for the purposes of the present deductions, these relations are

$$(21)\quad \begin{cases} P_\nu^{-\mu}(z) = (2\pi)^{-\frac{1}{2}} \dfrac{\Gamma(\mu + \frac{1}{2})}{\Gamma(2\mu + 1)} (t^2 - 1)^{\frac{1}{4}} \left(\dfrac{t+1}{t-1}\right)^{-\frac{1}{2}(\nu+\frac{1}{2})} Y_{\mu-\frac{1}{2},2}^{(\nu+\frac{1}{2},-\nu-\frac{1}{2})}(t), \\ Q_\nu^\mu(z) = (\frac{1}{2}\pi)^{\frac{1}{2}} e^{\mu\pi i} \dfrac{\Gamma(\nu + \mu + 1)}{\Gamma(\nu+\frac{3}{2})} (t^2 - 1)^{\frac{1}{4}} \left(\dfrac{t+1}{t-1}\right)^{-\frac{1}{2}(\nu+\frac{1}{2})} Y_{\mu-\frac{1}{2},1}^{(\nu+\frac{1}{2},-\nu-\frac{1}{2})}(t), \end{cases}$$

in which the symbols $Y_{\mu-\frac{1}{2},j}^{(\nu+\frac{1}{2},-\nu-\frac{1}{2})}(t)$ are specializations of the Jacobi functions (2). The relations (21) together with the well known identity

$$(22)\qquad P_\nu^\mu(z) = \frac{\Gamma(\nu + \mu + 1)}{\Gamma(\nu - \mu + 1)} P_\nu^{-\mu}(z) + \frac{2}{\pi} e^{-\mu\pi i} \sin \mu\pi\, Q_\nu^\mu(z)$$

will furnish complete asymptotic representations for the Legendre functions $P_\nu^\mu(z)$ and $Q_\nu^\mu(z)$ when $|\mu|$ is large and $|\nu|$ is bounded.

In substituting the asymptotic forms for the Jacobi functions into the relations (21) the variables Φ and ξ will be replaced by the quantities

$$(23)\qquad \bar{\Phi} = \log \left(\frac{z+1}{z-1}\right)^{\frac{1}{2}}, \qquad \bar{\xi} = i\mu\bar{\Phi};$$

in the first of these formulas the root is to be chosen as that which has a real positive value when z is real and greater than unity, and the logarithm is to have its principal value.

Substituting (14) into (21), one obtains the representations

$$(24)\quad \begin{cases} P_\nu^{-\mu}(z) = 2^{2\mu-\frac{1}{2}} e^{(\nu+\frac{1}{2})\frac{1}{2}\pi i} \dfrac{\mu^{\frac{1}{2}}\Gamma(\mu+\frac{1}{2})}{\Gamma(2\mu+1)} \bar{\Phi}^{\frac{1}{2}}\{H^{(1)}_{\nu+\frac{1}{2}}(i\mu\bar{\Phi}) + R_1\}, \\ Q_\nu^\mu(z) = \left(\dfrac{\pi}{2}\right)^{\frac{1}{2}} e^{(\mu-\frac{1}{2}\nu-\frac{1}{4})\pi i} \mu^{-\nu-\frac{1}{2}} \Gamma(\mu+\nu+1)\bar{\Phi}^{\frac{1}{2}}\{J_{\nu+\frac{1}{2}}(i\mu\bar{\Phi}) + R_2\}, \end{cases}$$

in which the correction terms R_i $(i = 1, 2)$ have the structural forms

$$R_1 = \begin{cases} \bar{\Phi}^{-\nu-\frac{1}{2}} O(\mu^{-\nu-\frac{3}{2}}) & \text{if } \nu + \frac{1}{2} \neq 0, \text{ for } |z| \geqq c|\mu|, \\ (\log \mu\bar{\Phi}) O(\mu^{-1}) & \text{if } \nu + \frac{1}{2} = 0, \text{ for } |z| \geqq c|\mu|, \\ \bar{\Phi}^{-\frac{1}{2}}\{e^{\mu\bar{\Phi}} O(\mu^{-\frac{1}{2}}) + e^{-\mu\bar{\Phi}} O(\mu^{-\frac{1}{2}})\} & \text{for } |z| < c|\mu|, \end{cases}$$

and

$$R_2 = \begin{cases} \bar{\Phi}^{\nu+\frac{3}{2}} O(\mu^{\nu+\frac{1}{2}}) & \text{for } |z| \geqq c|\mu|, \\ R_1 & \text{for } |z| < c|\mu|. \end{cases}$$

The symbol c denotes some positive constant and the quantity $\bar{\Phi}$ is given explicitly in (23).

If the identity (22) is used, it is convenient to restrict the parameter μ by the relation $|\arg \mu| \leqq \pi - \delta < \pi$ in order that use may be made of the asymptotic formula for the gamma function. In virtue of the results obtained for the function $P_\nu^{-\mu}(z)$ such a restriction involves no loss of generality. One obtains the representation

$$(25)\quad \begin{aligned} P_\nu^\mu(z) = e^{-\mu}\mu^\mu\bar{\Phi}^{\frac{1}{2}}\{(\sin \nu\pi) e^{-(\mu-\frac{1}{2}\nu+\frac{1}{4})\pi i} H^{(1)}_{\nu+\frac{1}{2}}(i\mu\bar{\Phi}) \\ + (\sin \mu\pi) e^{-(\nu+\frac{1}{2})\frac{1}{2}\pi i} H^{(2)}_{\nu+\frac{1}{2}}(i\mu\bar{\Phi}) + R_3\} \end{aligned}$$

in which the correction term R_3 has the forms

$$R_3 = \begin{cases} \sin(\mu-\nu)\pi\, O(\mu^{-1}) + \sin \mu\pi\, O(\mu^{-2}) & \text{if } \nu + \frac{1}{2} \neq 0, \text{ for } |z| \geqq c|\mu|, \\ \sin(\mu-\nu)\pi \log[1 + O(\mu^{-1})] O(\mu^{-1}) + \sin \mu\pi\, O(\mu^{-2}) & \\ \qquad\qquad & \text{if } \nu + \frac{1}{2} = 0, \text{ for } |z| \geqq c|\mu|, \\ \bar{\Phi}^{-\frac{1}{2}} O(\mu^{-\frac{1}{2}})\{\sin(\mu-\nu)\pi\, e^{-\mu\bar{\Phi}} + \sin \mu\pi\, e^{\mu\bar{\Phi}}\} & \text{for } |z| < c|\mu|. \end{cases}$$

A similar treatment involving the use of the asymptotic forms (15) in the relations (21) leads to representations in terms of elementary functions which are valid when $|z| < c|\mu|$. The results are

$$(26)\quad \begin{cases} P_\nu^{-\mu}(z) = \dfrac{2^{2\mu}}{\pi^{\frac{1}{2}}} \dfrac{\Gamma(\mu+\frac{1}{2})}{\Gamma(2\mu+1)} \left\{ c^{(h)}_{1,1} \left(\dfrac{z-1}{z+1}\right)^{\frac{1}{2}\mu} + c^{(h)}_{1,2} \left(\dfrac{z+1}{z-1}\right)^{\frac{1}{2}\mu} \right\}, \\ Q_\nu^\mu(z) = \frac{1}{2} e^{\mu\pi i} \mu^{-\nu-1} \Gamma(\nu+\mu+1) \left\{ c^{(h)}_{2,1} \left(\dfrac{z-1}{z+1}\right)^{\frac{1}{2}\mu} + c^{(h)}_{2,2} \left(\dfrac{z+1}{z-1}\right)^{\frac{1}{2}\mu} \right\}, \end{cases}$$

in which the coefficients $c^{(h)}_{i,j}$ are as shown in Table II.

TABLE II

h	$c_{1,1}^{(h)}$	$c_{1,2}^{(h)}$	$c_{2,1}^{(h)}$	$c_{2,2}^{(h)}$
1	[1]	0	$[-e^{\nu\pi i}]$	[1]
0	[1]	0	$[-e^{-\nu\pi i}]$	[1]
-1	[1]	$[2i\cdot\sin\nu\pi]$	$[-e^{-\nu\pi i}]$	$[e^{-2\nu\pi i}]$

When the forms (26) are substituted into the identity (22), it should be noted that the relation $|\arg\mu| \leqq \pi - \delta < \delta$ allows the restriction of the variable $\bar{\xi}$ to the pair of regions $\Xi^{(h)}$, for $h = 0, 1$. Easy computations lead to the formula

$$(27)\qquad P_\nu^\mu(z) = \left(\frac{2}{\pi}\right)^{\frac{1}{2}} e^{-\mu}\mu^{\mu-\frac{1}{2}}\left\{[\sin\mu\pi]\left(\frac{z+1}{z-1}\right)^{\frac{1}{2}\mu} - [(\sin\nu\pi)e^{\pm\mu\pi i}]\left(\frac{z-1}{z+1}\right)^{\frac{1}{2}\mu}\right\}$$

in which the upper or the lower sign is to be taken according as the variable $\bar{\xi}$ is in the region $\Xi^{(1)}$ or $\Xi^{(0)}$.

THEOREM III. *The representations, asymptotic as to μ, for the generalized Legendre functions $P_\nu^{\pm\mu}(z)$ and $Q_\nu^\mu(z)$ are given for general values of z by* (24) *and* (25). *For values of z such that $|z| < c|\mu|$ the formulas* (26) *and* (27) *apply. The forms are uniform with respect to ν for ν in any fixed bounded domain. The variable z ranges over the domain specified by $R(z) \geqq 0$ and $|\arg(z \pm 1)| \leqq \pi$. The range of* $\arg\mu$ *for* (24) *and* (26) *is given by* $-\pi - \delta < \arg\mu \leqq \pi - \delta$ *and for* (25) *and* (27) *by* $|\arg\mu| \leqq \pi - \delta < \pi$.

Correct forms for $P_\nu^\mu(\cos\theta)$ and $Q_\nu^\mu(\cos\theta)$, asymptotic as to μ, apparently have never been given. Barnes [1] published erroneous results. It is found from (27) and the second formula of (26) that the forms

$$P_\nu^\mu(\cos\theta) = \left(\frac{2}{\pi}\right)^{\frac{1}{2}} e^{-\mu}\mu^{\mu-\frac{1}{2}}\{[\sin\mu\pi](\cot\tfrac{1}{2}\theta)^\mu - [\sin\nu\pi](\tan\tfrac{1}{2}\theta)^\mu\},$$

$$Q_\nu^\mu(\cos\theta) = \left(\frac{\pi}{2}\right)^{\frac{1}{2}} e^{-\mu}\mu^{\mu-\frac{1}{2}}\{[\cos\mu\pi](\cot\tfrac{1}{2}\theta)^\mu - [\cos\nu\pi](\tan\tfrac{1}{2}\theta)^\mu\}$$

are valid when $|\mu|$ is large in comparison with $|\nu|$ and $|\arg\mu| \leqq \pi - \delta$, $0 < \theta \leqq \frac{1}{2}\pi$.

A few remarks concerning the regions $\Xi^{(h)}$ of validity for the forms (26) and (27) may be of assistance to the reader. The regions are given by (11) with $\bar{\xi}$ replacing ξ. The domain $R_{\bar{\Phi}}$ is described by the inequalities: $-\frac{1}{2}\pi \leqq I(\bar{\Phi}) < \frac{1}{2}\pi$, $-\frac{1}{2}\pi \leqq \arg\bar{\Phi} < \frac{1}{2}\pi$. Since $\arg\bar{\xi} = \frac{1}{2}\pi + \arg\mu + \arg\bar{\Phi}$, it is seen that, for a fixed value of $\arg\mu$, any boundary line of a region $\Xi^{(h)}$ which appears explicitly in $R_{\bar{\Phi}}$ will be a radial line extending from the origin at some angle of inclination α, $-\frac{1}{2}\pi \leqq \alpha \leqq \frac{1}{2}\pi$, with the axis of reals. The corresponding

curve in the z-plane will depart from a point on the real axis between 0 and 1 and will have the inclination $-\alpha$. As $|z|$ increases, the boundary line in question bends toward and finally becomes asymptotic to the line $\arg(z - 1) = -\alpha$. With these few facts, the reader may easily translate the discussion of §4 to suit the case in hand.

Bibliography

1. E. W. Barnes, *On the generalized Legendre functions*, Quar. Jour. of Math., vol. 39(1908), p. 97.
2. G. Darboux, *Sur l'approximation des fonctions*, etc., Liouv. Jour., (3), vol. 4(1878), p. 377.
3. E. W. Hobson, *On a general type of spherical harmonic*, etc., Phil. Trans. Royal Soc. of London, vol. 97(1896).
4. E. W. Hobson, *Spherical and Ellipsoidal Harmonics*, Cambridge, England, 1931.
5. C. G. J. Jacobi, *Untersuchungen über die Differentialgleichung der hypergeometrischen Reihe*, Werke, vol. 6, p. 184, or Jour. für Math., vol. 56(1859), p. 149.
6. R. E. Langer, *On the asymptotic solutions of ordinary differential equations*, etc., Trans. Amer. Math. Soc., vol. 37(1935), p. 397.
7. G. Szegö, *Asymptotische Entwicklungen der Jacobischen Polynome*, Halle, 1933.
8. G. N. Watson, *Asymptotic expansions of hypergeometric functions*, Trans. Cambridge Phil. Soc., vol. 22.
9. F. J. W. Whipple, *A symmetric relation between Legendre's functions*, Proc. London Math. Soc., (2), vol. 16(1917).

Ohio State University.

PRESERVATION OF PARTIAL LIMITS IN MULTIPLE SEQUENCE TRANSFORMATIONS

By Hugh J. Hamilton

1. **Introduction.** Problems (iii) and (iv) in §1.4 of[1] H_1 were solved in that paper only for $s_{k^2} = 0$. It is the purpose of the present paper to give complete solutions. Existence of the transform σ_m is assumed for each m.

2. **Additional notations and definitions.** Let X and Y denote classes of sequences. The notation X $\rightarrow$ Y *row reg* shall signify row regularity of the transformation X $\rightarrow$ Y in case each of X and Y is the class of regularly convergent sequences, and ultimate row regularity in all other cases in question. (See §1.4 of H_1.) Thus[2] NS RC $\rightarrow$ RC *row reg* will mean NS RC $\rightarrow$ RC with $\sigma_{k^2} = s_{k^2}$ for all k^2. The notation NS RC $\rightarrow$ RC *ul row reg* shall mean NS RC $\rightarrow$ RC with $\sigma_{k^2} = s_{k^2}$ for all k^2 sufficiently large.

Consider the matrix $\| b_{mk} \|$, where $b_{mk} = a_{mk}$ $(k \neq m)$, and $b_{mm} = a_{mm} - 1$. Define the sequence $\{\tau_m\}$ by the equations

$$\tau_m = \sum_{k=1}^{\infty} b_{mk} s_k \equiv \sigma_m - s_m,$$

and let NS X-* $\rightarrow$ Y denote NS $\{\tau_m\}$ be of class Y whenever $\{s_k\}$ is of class X.

3. **List of theorems (first form).** The following theorems are obvious.

NS RC, NS BURC, NS URC $\rightarrow$ URC row reg are, respectively, NS RC, NS BURC, NS URC- $\rightarrow$ URCRN.*

NS RC, NS BURC $\rightarrow$ BURC row reg are, respectively, NS RC, NS BURC- $\rightarrow$ BURCRN.*

NS URC $\rightarrow$ BURC row reg are NS URC $\rightarrow$ B and NS URC- $\rightarrow$ URCRN.*

NS RC $\rightarrow$ RC ul row reg are NS RC- $\rightarrow$ RCURN.*

NS BURC, NS URC $\rightarrow$ RC row reg are, respectively, NS BURC, NS URC $\rightarrow$ RC, and, respectively, NS BURC, NS URC- $\rightarrow$ URCRN.*

NS RC $\rightarrow$ RC row reg are NS RC- $\rightarrow$ RCRN.*

Received October 28, 1938.

[1] H_1 will denote the author's paper, *Transformations of multiple sequences*, this Journal, vol. 2(1936), pp. 29–60. The present paper assumes familiarity with the contents of H_1, the ideas, terminology, notations, and results of which are freely used without further comment.

[2] NS shall abbreviate *conditions necessary and sufficient that.*

4. **List of theorems (second form).** Denote each condition listed in §3 of H_1, as applied to the matrix $\| b_{mk} \|$, by the label there attached to it, with a prime. (Thus $(a_1)'$, $(a_2)'$, etc.) It is then possible to express the theorems of §3 in a fashion typified by the following two particular examples.

NS RC → URC row reg are $(a_1)'$, $(b_1)'$, $(d_1)'$, $(d_2)'$, $(d_3)'$, $(\bar{e}_1)'$, $(\bar{e}_2)'$, $(\bar{e}_3)'$.

NS URC → BURC row reg are (c_1), (c_2), $(a_1)'$, $(a_2)'$, $(b_1)'$, $(b_2)'$, $(d_1)'$, $(d_3)'$, $(\bar{e}_1)'$, $(\bar{e}_2)'$, $(\bar{e}_3)'$.

5. **Reductions of conditions.** These sets of conditions may be more elegantly expressed by reducing the conditions $(a_1)'$, $(a_2)'$, etc., to simpler conditions on the matrix $\| a_{mk} \|$. For the most part the latter occur among those listed in §3 of H_1. However, the following are also needed.

$\{\bar{e}_2\}$ (e_2), with
$$L_{m^1k^3} = 0 \text{ for } k^3 \neq m^1,\ k^3 \neq m^{11};$$
$$L_{m^1k^3} = 1 \text{ for } k^3 = m^1 \text{ or } k^3 = m^{11}.$$

$\{\bar{e}_3\}$ (e_3), with $L_{m^1} = 1$.

$\{\bar{f}_2\}$ (f_2), with
$$L_{m^1k^3} = 0 \text{ for } k^3 \neq m^1,\ k^3 \neq m^{11};$$
$$L_{m^1k^3} = 1 \text{ for } k^3 = m^1 \text{ or } k^3 = m^{11}.$$

$\{\bar{f}_3\}$ (f_3), with $L_{m^1} = 1$.

In the list of equivalences, which follows, note that the conditions in the first paragraph involve the mere *existence* of limits, bounds, etc.

$(a_1)'$, $(a_2)'$, $(b_1)'$, $(b_2)'$, $(c_1)'$, $(d_1)'$, $(d_2)'$, $(d_3)'$, $(d_4)'$, $(e_4^*)'$, $(\bar{e}_1)'$, $(f_1)'$, $(f_2)'$, $(f_3)'$, $(\bar{f}_1)'$ are equivalent, respectively, to (a_1), (a_2), (b_1), (b_2), (c_1), (d_1), (d_2), (d_3), (d_4), (e_4^*), $(\bar{e}_1)$, (f_1), (f_2), (f_3), $(\bar{f}_1)$.

$(\bar{e}_2)'$, $(\bar{e}_3)'$, $(\bar{f}_2)'$, $(\bar{f}_3)'$ are equivalent, respectively, to $\{\bar{e}_2\}$, $\{\bar{e}_3\}$, $\{\bar{f}_2\}$, $\{\bar{f}_3\}$.

The equivalences in the first paragraph are obvious. Proof will now be given for the first equivalence in the second paragraph.

The condition

$$\lim_{m^2=\infty} \sum_{k^4=1}^{\infty} b_{mk} = 0 \qquad \text{for } m^1 > \bar{E} \qquad (k^3 = 1, 2, \cdots)$$

becomes

$$\lim_{m^2=\infty} \sum_{k^4=1}^{\infty} a_{mk} = 0 \qquad \text{for } m^1 > \bar{E} \qquad (k^3 = 1, 2, \cdots),$$

excepting cases in which the positions of the elements of k^3 occur among those of m^1 and corresponding values coincide, and in these cases becomes

$$\lim_{m^2=\infty} \sum_{k^4=1}^{\infty} a_{mk} = 1 \qquad \text{for } m^1 > \bar{E}.$$

6. **List of theorems (final form).** Finally, in view of (.08)–(.11) of §4 in H_1, the sought results may be tabulated as follows.

1. *NS RC → URC row reg are* (a_1), (b_1), (d_1), (d_2), (d_3), ($\bar{e}_1$), $\{\bar{e}_2\}$, $\{\bar{e}_3\}$.

2. *NS BURC → URC row reg are* (a_1), (b_1), (d_3), (d_4), (e_4^*), ($\bar{e}_1$), $\{\bar{e}_2\}$, $\{\bar{e}_3\}$.

3. *NS URC → URC row reg are* (a_1), (a_2), (b_1), (b_2), (d_1), (d_3), ($\bar{e}_1$), $\{\bar{e}_2\}$, $\{\bar{e}_3\}$.

4. *NS RC → BURC row reg are* (c_1), (d_1), (d_2), (d_3), ($\bar{e}_1$), $\{\bar{e}_2\}$, $\{\bar{e}_3\}$.

5. *NS BURC → BURC row reg are* (c_1), (d_3), (d_4), (e_4^*), ($\bar{e}_1$), $\{\bar{e}_2\}$, $\{\bar{e}_3\}$.

6. *NS URC → BURC row reg are* (c_1), (c_2), (d_1), (d_3), ($\bar{e}_1$), $\{\bar{e}_2\}$, $\{\bar{e}_3\}$.

7. *NS RC → RC ul row reg are* (c_1), (d_1), (d_2), (d_3), ($\bar{e}_1$), $\{\bar{e}_2\}$, $\{\bar{e}_3\}$, (f_1), (f_2), (f_3).

8. *NS BURC → RC row reg are* (c_1), (d_3), (d_4), (e_4^*), ($\bar{e}_1$), $\{\bar{e}_2\}$, $\{\bar{e}_3\}$, (f_3), (f_4).

9. *NS URC → RC row reg are* (c_1), (c_2), (d_1), (d_3), ($\bar{e}_1$), $\{\bar{e}_2\}$, $\{\bar{e}_3\}$, (f_1), (f_3).

10. *NS RC → RC row reg are* (c_1), (d_1), (d_2), (d_3), ($\bar{f}_1$), $\{\bar{f}_2\}$, $\{\bar{f}_3\}$.

7. **A verification of sufficiency.** Checks for these results are available in the formulas for row limits given in H_1. Thus, for example, formula (46.1), p. 51, of H_1, and 1 above yield, for r^1 sufficiently large,

$$\sigma_{r^1} - s = \sum{}' (s_{k^3} - s)\left\{1 - \sum_{\tau=1}^{T-\rho} (-1)^{\tau+1} \sum{}'' 1\right\},$$

where T is the dimension of r^1, ρ is that of k^3, and τ is that of k^{41}; $\sum'$ sums over all k^3 whose elements are included among those of r^1, and $\sum''$ sums over all k^{41} whose elements are likewise among those of r^1. The part of the expansion of the right side corresponding to $k^3 = r^1$ is $s_{r^1} - s$. That corresponding to any other k^3 is

$$(s_{k^3} - s)\left\{1 - \sum_{\tau=1}^{T-\rho} (-1)^{\tau+1}\binom{T-\rho}{\tau}\right\} = (s_{k^3} - s)\sum_{\tau=0}^{T-\rho} (-1)^{\tau}\binom{T-\rho}{\tau} = 0.$$

Hence $\sigma_{r^1} = s_{r^1}$.

8. **Application to double sequences.** As application, 10 yields the following set of conditions on the four-dimensional matrix $\| a_{pqij} \|$ necessary and sufficient that the double sequence $\{\sigma_{pq}\}$ be rc with all row and column limits the same as those of $\{s_{ij}\}$ whenever $\{s_{ij}\}$ is rc, where $\sigma_{pq} = \sum_{i,j=1}^{\infty} a_{pqij}s_{ij}$.

NS RC → RC row reg are:

(c_1) $$\sum_{i,j=1}^{\infty} |a_{pqij}| < A \qquad (p, q = 1, 2, \cdots);$$

$$(\mathrm{d}_1) \qquad \lim_{p,q=\infty} a_{pqij} = a_{ij} \qquad (i, j = 1, 2, \cdots);$$

$$(\mathrm{d}_2) \qquad \begin{aligned} &\lim_{p,q=\infty} \sum_{i=1}^{\infty} a_{pqij} = L_j && (j = 1, 2, \cdots), \\ &\lim_{p,q=\infty} \sum_{j=1}^{\infty} a_{pqij} = L_i' && (i = 1, 2, \cdots); \end{aligned}$$

$$(\mathrm{d}_3) \qquad \lim_{p,q=\infty} \sum_{i,j=1}^{\infty} a_{pqij} = L;$$

$$(\bar{\mathrm{f}}_1) \qquad \begin{aligned} &\lim_{p=\infty} a_{pqij} = 0 \quad \textit{for all } q && (i, j = 1, 2, \cdots), \\ &\lim_{q=\infty} a_{pqij} = 0 \quad \textit{for all } p && (i, j = 1, 2, \cdots); \end{aligned}$$

$$\{\bar{\mathrm{f}}_2\} \qquad \begin{aligned} &\lim_{p=\infty} \sum_{i=1}^{\infty} a_{pqij} = 0 \quad \textit{for all } q \\ &\qquad\qquad (j = 1, 2, \cdots, q-1, q+1, q+2, \cdots), \\ &\lim_{p=\infty} \sum_{i=1}^{\infty} a_{pqiq} = 1 \quad \textit{for all } q, \\ &\lim_{p=\infty} \sum_{j=1}^{\infty} a_{pqij} = 0 \quad \textit{for all } q \qquad (i = 1, 2, \cdots), \\ &\lim_{q=\infty} \sum_{i=1}^{\infty} a_{pqij} = 0 \quad \textit{for all } p \qquad (j = 1, 2, \cdots), \\ &\lim_{q=\infty} \sum_{j=1}^{\infty} a_{pqij} = 0 \quad \textit{for all } p \\ &\qquad\qquad (i = 1, 2, \cdots, p-1, p+1, p+2, \cdots), \\ &\lim_{q=\infty} \sum_{j=1}^{\infty} a_{pqpj} = 1 \quad \textit{for all } p; \end{aligned}$$

$$\{\bar{\mathrm{f}}_3\} \qquad \begin{aligned} &\lim_{p=\infty} \sum_{i,j=1}^{\infty} a_{pqij} = 1 \quad \textit{for all } q, \\ &\lim_{q=\infty} \sum_{i,j=1}^{\infty} a_{pqij} = 1 \quad \textit{for all } p. \end{aligned}$$

9. **Implications of the new conditions**. Finally, we may note certain implications of conditions $\{\bar{\mathrm{e}}_2\}$, $\{\bar{\mathrm{e}}_3\}$, $\{\bar{\mathrm{f}}_2\}$, and $\{\bar{\mathrm{f}}_3\}$, to wit: $\{\bar{\mathrm{e}}_2\} \rightarrow (\mathrm{e}_2)$, $\{\bar{\mathrm{e}}_3\} \rightarrow (\mathrm{e}_3)$, $\{\bar{\mathrm{f}}_2\} \rightarrow (\mathrm{f}_2)$, $\{\bar{\mathrm{f}}_3\} \rightarrow (\mathrm{f}_3)$, $\{\bar{\mathrm{f}}_2\} \rightarrow \{\bar{\mathrm{e}}_2\}$, $\{\bar{\mathrm{f}}_3\} \rightarrow \{\bar{\mathrm{e}}_3\}$, $(\mathrm{d}_2) + \{\bar{\mathrm{e}}_2\} \rightarrow (\bar{\mathrm{d}}_2)$, $(\mathrm{d}_3) + \{\bar{\mathrm{e}}_3\}$

$\rightarrow$ (d_3) with $L = 1$. (Compare .75, p. 41, of H_1.) With these relations and those on pp. 37–41 of H_1, the sets of conditions necessary and sufficient for the various transformations can be materially strengthened. For example, NS RC $\rightarrow$ RC row reg may be written: (c_1), (d_1), (d_2), (d_3) with $L = 1$, ($\hat{f}_1$), $\{\hat{f}_2\}$, $\{\hat{f}_3\}$. Thus in the above application to double sequences a_{ij}, L_j, and L_i' may be replaced by 0 ($i, j = 1, 2, \cdots$), and L by 1.

POMONA COLLEGE.

CONVERGENCE THEOREMS FOR CONTINUED FRACTIONS

BY WALTER LEIGHTON

1. **Introduction**. The purpose of this paper is to present a new set of convergence theorems for continued fractions of the form

$$(1.1)\qquad 1 + \frac{a_1}{1+}\,\frac{a_2}{1+}\,\frac{a_3}{1+}\cdots,$$

where the a_n are complex numbers $\neq 0$. The method used is an extension of a method used in an earlier paper (Leighton [1][1]) the results of which now follow from Theorem 4.1 of the present paper.

A number of writers[2] proved independently that if $|a_n| \leqq \frac{1}{4}$ $(n = 2, 3, 4, \cdots)$, the continued fraction (1.1) converges. Szász [1] showed that the constant $\frac{1}{4}$ cannot be improved by proving that the continued fraction

$$\frac{-\frac{1}{4}-e}{1}\;+\;\frac{-\frac{1}{4}-e}{1}\;+\;\frac{-\frac{1}{4}-e}{1}\;+\cdots$$

diverges for each value of $e > 0$. Later, new types of sufficient conditions for convergence were found (Leighton and Wall [1], Jordan and Leighton [1], Leighton [2]), but all of these theorems required that at least an infinite subsequence of the $|a_n| \leqq \frac{1}{4}$. This last condition was recently removed (Leighton [1]) by showing that (1.1) converges if

$$(1.1)'\qquad |1 + a_2| \geqq 1 + |a_1|, \qquad |a_2| \geqq \frac{2+m}{1-m},$$
$$|a_{2n+1}| \leqq m < 1, \qquad |a_{2n+2}| \geqq 2 + m + m\,|a_{2n}| \qquad (n = 1, 2, 3, \cdots)$$

It will follow incidentally from Theorem 4.4 of the present paper that this condition can be removed in still different ways.

We recall that the n-th approximant A_n/B_n of a continued fraction

$$(1.2)\qquad \beta_0 + \frac{\alpha_1}{\beta_1+}\,\frac{\alpha_2}{\beta_2+}\cdots$$

is defined by means of the recursion relations

$$(1.3)\qquad \begin{aligned} &A_0 = \beta_0, \qquad B_0 = 1, \qquad A_1 = \beta_0\beta_1 + \alpha_1, \qquad B_1 = \beta_1, \\ &A_n = \beta_n A_{n-1} + \alpha_n A_{n-2}, \qquad B_n = \beta_n B_{n-1} + \alpha_n B_{n-2}, \quad (n = 2, 3, 4, \cdots). \end{aligned}$$

Received November 15, 1938.

[1] Numbers in brackets refer to the bibliography.

[2] For bibliography on this criterion see Szász [1] and Leighton [1].

It will be useful to designate by $A_{n,\lambda}/B_{n,\lambda}$ (Perron [1], p. 14) the n-th approximant of the truncated continued fraction

$$\beta_\lambda + \frac{\alpha_{\lambda+1}}{\beta_{\lambda+1} +} \frac{\alpha_{\lambda+2}}{\beta_{\lambda+2} +} \cdots . \tag{1.4}$$

The plan of the present paper is to form the continued fractions

$$d_0 + \frac{c_1}{d_1 +} \frac{c_2}{d_2 +} \cdots , \tag{1.5}$$

the approximants C_n/D_n of which are related to the approximants A_n/B_n of (1.1) as follows:

$$\begin{aligned} C_n &= A_{kn+r}, \qquad & D_n &= B_{kn+r}, \qquad & (n &= 1, 2, 3, \cdots), \\ C_0 &= \frac{A_r}{B_r}, & D_0 &= 1, \end{aligned} \tag{1.6}$$

where k is a *fixed* positive integer $\geqq 2$ and r is chosen from the numbers $0, 1, 2, \cdots, k-1$. We are thus led to a set of k continued fractions (1.5) with the property that the k sequences of approximants C_n/B_n generated by them together comprise the totality of approximants A_n/B_n of the given continued fraction (1.1). A condition of Pringsheim (Perron [1], p. 254) for convergence will be applied to (1.5). Thus, each of the k continued fractions will converge. An additional condition will be added to insure that these continued fractions converge to the same limit. It will follow that (1.1) converges under these conditions. The result will be the sets of convergence criteria referred to in the first paragraph.

2. **A lemma**. It is well known (Perron [1], p. 198) that if $t_0, t_1, t_2, \cdots$ is any sequence of complex numbers, where $t_{n-1} \neq t_n$ $(n = 1, 2, 3, \cdots)$, the continued fraction (1.2) with

$$\begin{aligned} \beta_0 &= t_0, \\ \alpha_1 &= t_1 - t_0, \qquad & \beta_1 &= 1, \\ \alpha_n &= \frac{t_{n-1} - t_n}{t_{n-1} - t_{n-2}}, \qquad & \beta_n &= \frac{t_n - t_{n-2}}{t_{n-1} - t_{n-2}}, \qquad (n = 2, 3, 4, \cdots) \end{aligned} \tag{2.1}$$

has approximants A_n/B_n with the property that

$$A_n = t_n, \qquad B_n = 1, \qquad (n = 0, 1, 2, \cdots).$$

If the numbers t_n are defined as quotients r_n/s_n, it will be useful to determine numbers c_n and d_n such that the n-th approximant C_n/D_n of a continued fraction of the form (1.5) shall satisfy the following relations:

$$\begin{aligned} C_n &= r_n, \qquad & D_n &= s_n, \qquad & (n &= 1, 2, 3, \cdots), \\ C_0 &= \frac{r_0}{s_0}, & D_0 &= 1. \end{aligned} \tag{2.2}$$

To this end we shall recall (Perron [1], p. 196) that the n-th approximant U_n/V_n of the continued fraction

$$\beta_0 + \frac{u_1\alpha_1}{u_1\beta_1 +} \frac{u_1u_2\alpha_2}{u_2\beta_2 +} \frac{u_2u_3\alpha_3}{u_3\beta_3 +} \cdots + \frac{u_{n-1}u_n\alpha_n}{u_n\beta_n +} \cdots$$

has the properties

$$(2.3)\qquad \begin{aligned} &U_0 = \frac{r_0}{s_0}, \qquad U_1 = u_1\frac{r_1}{s_1}, \qquad V_1 = u_1, \\ &U_n = u_1u_2\cdots u_n\frac{r_n}{s_n}, \qquad V_n = u_1u_2\cdots u_n, \qquad (n = 2, 3, 4, \cdots). \end{aligned}$$

The proof of the following lemma is now immediate.

LEMMA. *If the numbers c_n and d_n are defined by the equations*

$$(2.4)\qquad \begin{aligned} &d_0 = \frac{r_0}{s_0}, \qquad c_1 = s_1\left[\frac{r_1}{s_1} - \frac{r_0}{s_0}\right], \qquad d_1 = s_1, \\ &c_n = \frac{s_n}{s_{n-2}}\cdot\left[\frac{\dfrac{r_{n-1}}{s_{n-1}} - \dfrac{r_n}{s_n}}{\dfrac{r_{n-1}}{s_{n-1}} - \dfrac{r_{n-2}}{s_{n-2}}}\right], \\ &d_n = \frac{s_n}{s_{n-1}}\cdot\left[\frac{\dfrac{r_n}{s_n} - \dfrac{r_{n-2}}{s_{n-2}}}{\dfrac{r_{n-1}}{s_{n-1}} - \dfrac{r_{n-2}}{s_{n-2}}}\right], \qquad (n = 2, 3, 4, \cdots), \end{aligned}$$

the n-th approximant C_n/D_n of the continued fraction

$$(2.5)\qquad d_0 + \frac{c_1}{d_1 +} \frac{c_2}{d_2 +} \cdots$$

has the properties

$$(2.6)\qquad \begin{aligned} &C_n = r_n, \qquad D_n = s_n, \qquad (n = 1, 2, 3, \cdots), \\ &C_0 = \frac{r_0}{s_0}, \qquad D_0 = 1. \end{aligned}$$

It is sufficient to observe that in (2.3) one may set $u_1 = s_1$, $u_n = s_n/s_{n-1}$ $(n = 2, 3, 4, \cdots)$.

3. **The principal theorem**. In the following result m is any fixed positive constant < 1 and r_0 is one of the integers $0, 1, 2, \cdots, k - 1$. The integer r_0, once chosen, is to be kept fixed, but its choice is quite arbitrary. The integer k is fixed $\geqq 2$.

THEOREM 3.1. *If the denominators B_r $(r = 0, 1, 2, \cdots, k-1)$ of the first k approximants of the continued fraction*

$$1 + \frac{a_1}{1+}\,\frac{a_2}{1+}\cdots \tag{3.1}$$

are $\neq 0$, if the numbers $B_{k-1,\lambda}$ $(\lambda = 1, 2, 3, \cdots)$ are $\neq 0$, if

$$|a_{nk+s+1}B_{k-2,nk+s+1}| \leqq m < 1 \tag{3.2}$$

for each value of s except those $\equiv r_0 \pmod{k}$, and if

$$|B_rB_{k+r}| \geqq |B_r| + |a_1a_2\cdots a_{r+1}B_{k-1,r+1}| \qquad (r = 0, 1, 2, \cdots, k-1), \tag{3.3}$$

$$|B_{2k-1,\lambda}| \geqq |B_{k-1,\lambda}| + |a_{\lambda+1}a_{\lambda+2}\cdots a_{\lambda+k}B_{k-1,\lambda+k}| \qquad (n = 1, 2, 3, \cdots), \tag{3.4}$$

the continued fraction (3.1) *converges.*

Thus for each integral value of k greater than or equal to 2, this theorem yields a convergence criterion corresponding to each of the k distinct choices of r_0.

To prove the theorem construct the continued fraction (2.5) the n-th approximant C_n/D_n of which satisfies the conditions

$$\begin{aligned} C_n &= A_{nk+r}, & C_0 &= \frac{A_r}{B_r}, \\ D_n &= B_{nk+r}, & D_0 &= 1. \end{aligned} \tag{3.5}$$

We find by means of formulas (2.4) that

$$d_0 = \frac{A_r}{B_r}, \qquad c_1 = B_{k+r}\left[\frac{A_{k+r}}{B_{k+r}} - \frac{A_r}{B_r}\right],$$

$$c_n = \frac{B_{nk+r}}{B_{(n-2)k+r}}\cdot\left[\frac{\dfrac{A_{(n-1)k+r}}{B_{(n-1)k+r}} - \dfrac{A_{nk+r}}{B_{nk+r}}}{\dfrac{A_{(n-1)k+r}}{B_{(n-1)k+r}} - \dfrac{A_{(n-2)k+r}}{B_{(n-2)k+r}}}\right],$$

$$d_n = \frac{B_{nk+r}}{B_{(n-1)k+r}}\cdot\left[\frac{\dfrac{A_{nk+r}}{B_{nk+r}} - \dfrac{A_{(n-2)k+r}}{B_{(n-2)k+r}}}{\dfrac{A_{(n-1)k+r}}{B_{(n-1)k+r}} - \dfrac{A_{(n-2)k+r}}{B_{(n-2)k+r}}}\right], \qquad (n = 2, 3, 4, \cdots).$$

One can simplify these equations by means of the well known formula (Perron [1], p. 17)

$$\frac{A_{n+\lambda-1}}{B_{n+\lambda-1}} - \frac{A_{\lambda-1}}{B_{\lambda-1}} = \frac{(-1)^{\lambda-1}a_1a_2\cdots a_\lambda B_{n-1,\lambda}}{B_{n+\lambda-1}B_{\lambda-1}}, \tag{3.6}$$

obtaining

$$
\begin{aligned}
&d_0 = \frac{A_r}{B_r}, \qquad c_1 = (-1)^r a_1 a_2 \cdots a_{r+1} \frac{B_{k-1,r+1}}{B_r}, \qquad d_1 = B_{k+r}, \\
&c_n = (-1)^{k+1} a_{(n-2)k+r+2} a_{(n-2)k+r+3} \cdots a_{(n-1)k+r+1} \frac{B_{k-1,(n-1)k+r+1}}{B_{k-1,(n-2)k+r+1}}, \\
&d_n = \frac{B_{2k-1,(n-2)k+r+1}}{B_{k-1,(n-2)k+r+1}}, \qquad (n = 2, 3, 4, \cdots).
\end{aligned}
\tag{3.7}
$$

It is clear that the numbers c_n and d_n are well-defined because of the precautionary hypotheses $B_r \neq 0$, $B_{k-1,\lambda} \neq 0$ of the theorem.

Pringsheim has shown (Perron [1], p. 254) that if the elements c_n and d_n of a continued fraction

$$d_0 + \frac{c_1}{d_1 +} \frac{c_2}{d_2 +} \cdots \tag{3.8}$$

satisfy the conditions

$$|d_n| \geqq |c_n| + 1 \qquad (n = 1, 2, 3, \cdots), \tag{3.9}$$

the continued fraction converges and the denominators of consecutive approximants have the property that

$$|D_n| - |D_{n-1}| \geqq |c_1 c_2 \cdots c_n| \qquad (n = 1, 2, 3, \cdots). \tag{3.10}$$

Conditions (3.9) subject to (3.7) are precisely conditions (3.3) together with

$$
\begin{aligned}
|B_{2k-1,(n-2)k+r+1}| &\geqq |B_{k-1,(n-2)k+r+1}| \\
&\quad + |a_{(n-2)k+r+2} a_{(n-2)k+r+3} \cdots a_{(n-1)k+r+1} B_{k-1,(n-1)k+r+1}|,
\end{aligned}
\tag{3.11}
$$

where $r = 0, 1, 2, \cdots, k-1$ and $n = 2, 3, 4, \cdots$. Conditions (3.11), however, are easily seen to be equivalent to conditions (3.4). Thus under the hypotheses of the theorem the continued fraction (3.8) subject to (3.7) converges for each of the k values of r. It follows that the sequences $\{A_{kn+r}/B_{kn+r}\}_{n=0}^{n=\infty}$, where A_{kn+r}/B_{kn+r} is the $(kn+r)$-th approximant of (3.1), converge for each $r = 0, 1, 2, \cdots, k-1$. It remains to prove that these k sequences have the same limit.

To that end we observe that it is sufficient to prove that $\lim_{n\to\infty} \Delta_n^r = 0$, where

$$\Delta_n^r \equiv \left| \frac{A_{kn+r+1}}{B_{kn+r+1}} - \frac{A_{kn+r}}{B_{kn+r}} \right| = \left| \frac{a_1 a_2 \cdots a_{kn+r+1}}{B_{kn+r} B_{kn+r+1}} \right|,$$

and $r = r_0, r_0 + 1, \cdots, r_0 + k - 2$. The latter equality is a consequence of (3.6). By (3.10) and (3.7) we have

$$|B_{(n+1)k+r}| - |B_{nk+r}| \geqq |a_1 a_2 \cdots a_{nk+r+1}| \cdot \left| \frac{B_{k-1,nk+r+1}}{B_r} \right|, \tag{3.12}$$

and hence

$$\Delta_n^r \leqq \frac{|B_r|}{|B_{nk+r+1} B_{k-1,nk+r+1}|} \cdot \left[\frac{|B_{(n+1)k+r}|}{|B_{nk+r}|} - 1 \right] \qquad (n = 0, 1, 2, \cdots). \tag{3.13}$$

By a well-known result (Perron [1], p. 18)

$$B_{nk+r+1}B_{k-1,nk+r+1} = B_{(n+1)k+r} - a_{nk+r+2}B_{k-2,nk+r+2}B_{nk+r}. \tag{3.14}$$

Thus

$$\Delta_n^r \leqq \left|\frac{B_r}{B_{nk+r}}\right| \cdot \left[\frac{1 - \dfrac{|B_{nk+r}|}{|B_{(n+1)k+r}|}}{\left|1 - a_{nk+r+2}B_{k-2,nk+r+2}\cdot\dfrac{B_{nk+r}}{B_{(n+1)k+r}}\right|}\right]. \tag{3.15}$$

Now (3.12) insures us that $|B_{nk+r}|$ is strictly increasing with n, and hence that $\lim_{n\to\infty}|B_{nk+r}| = B^{(r)}$, where $B^{(r)}$ may be infinite. Applying conditions (3.2) to (3.15), we observe that the denominator in the bracket is $>1 - m$. It follows that $\lim_{n\to\infty}\Delta_n^r = 0$ $(r = r_0, r_0 + 1, \cdots, r_0 + k - 2)$ whether $B^{(r)}$ is finite or infinite. The proof is complete.

If now in (3.11) we set $r = r_0, r_0 + 1, \cdots, r_0 + k - 2$ and apply conditions (3.2), we obtain the following result.

THEOREM 3.2. *If the denominators B_r $(r = 0, 1, 2, \cdots, k - 1)$ of the first k approximants of the continued fraction* (3.1) *are* $\neq 0$, *if the numbers $B_{k-1,\lambda}$ $(\lambda = 1, 2, 3, \cdots)$ are* $\neq 0$, *if*

$$|a_{nk+s+1}B_{k-2,nk+s+1}| \leqq m < 1, \tag{3.16}$$

for each value of $s \not\equiv r_0 \pmod{k}$, *and if*

$$|B_rB_{k+r}| \geqq |B_r| + |a_1a_2\cdots a_{r+1}B_{k-1,r+1}| \quad (r = 0, 1, 2, \cdots, k-1),$$

$$|B_{2k-1,\lambda}| \geqq \frac{m}{|a_\lambda|}[1 + |a_\lambda a_{\lambda+1}\cdots a_{\lambda+k-1}|] \tag{3.17}$$

$$(\lambda = \lambda_0 + 1, \lambda_0 + 2, \cdots, \lambda_0 + k - 1),$$

$$|B_{2k-1,\lambda_0}| \geqq |B_{k-1,\lambda_0}| + |a_{\lambda_0+1}a_{\lambda_0+2}\cdots a_{\lambda_0+k}B_{k-1,\lambda_0+k}| \tag{3.18}$$

where $n = 2, 3, \cdots$ *and* $\lambda_0 = (n-2)k + r_0 + 1$, *the continued fraction* (3.1) *converges.*

The proof is immediate. Corollaries of this theorem will be examined in the next section. We observe here that the theorem will remain valid if conditions (3.18) are replaced by the conditions

$$|B_{2k-1,\lambda_0}| \geqq |B_{k-1,\lambda_0}| + \frac{1}{m}|a_{\lambda_0+1}B_{2k-1,\lambda_0+1}B_{k-1,\lambda_0+k}|, \tag{3.19}$$

since (3.19) implies (3.18).

4. **Consequences of the preceding theorems.** In this section we shall derive a number of results from Theorems 3.1 and 3.2 by assigning special values to k.

Suppose $k = 2$. Then $B_r = 1$ $(r = 0, 1)$ and $B_{k-1,\lambda} = B_{1,\lambda} = 1$ $(\lambda =$

1, 2, 3, $\cdots$). We have the following corollary of Theorem 3.1, combining the case $r_0 = 0$ with that of $r_0 = 1$.

THEOREM 4.1. *If either* $|a_{2n+1}| \leqq m < 1$ $(n = 1, 2, 3, \cdots)$ *or* $|a_{2n}| \leqq m < 1$ $(n = 1, 2, 3, \cdots)$, *and if*

$$|1 + a_2| \geqq 1 + |a_1|,$$

$$|1 + a_n + a_{n+1}| \geqq 1 + |a_{n-1}a_n| \qquad (n = 2, 3, \cdots),$$

the continued fraction

$$(4.1) \qquad 1 + \frac{a_1}{1+}\frac{a_2}{1+}\cdots$$

converges.

These are particularly simple conditions. For a further discussion of the case $k = 2$ the reader is referred to an earlier paper (Leighton [1]). We observe that since (1.1)′ imply the conditions of Theorem 4.1, this theorem and hence Theorem 3.1 are independent of earlier criteria. It will follow from Theorem 4.4, for example, that Theorem 3.1 is indeed more general than Theorem 4.1 as well. It may be pointed out that Pringsheim's conditions (3.9) never apply to continued fractions of the form (4.1).

We now consider the case $k = 3$, taking $r_0 = 0$ for simplicity and observing that each theorem stated henceforth represents k theorems which can be derived from the given theorem by advancing suitably the subscripts involved. Here r runs over the values 0, 1, and 2, and $B_0 = B_1 = 1$, $B_2 = 1 + a_2$, $B_{k-1,\lambda} = B_{2,\lambda} = 1 + a_{\lambda+2}$. Referring to Theorem 3.2 we must require that

$$1 + a_n \neq 0 \qquad (n = 2, 3, \cdots),$$

$$(4.2) \qquad \begin{aligned} &|a_{3n+2}| \leqq m < 1, \qquad |a_{3n+3}| \leqq m < 1, \ (n = 0, 1, 2, \cdots),\\ &|B_3| \geqq 1 + |a_1(1 + a_3)|, \qquad |B_4| \geqq 1 + |a_1a_2(1 + a_4)|,\\ &|(1 + a_2)B_5| \geqq |1 + a_2| + |a_1a_2a_3(1 + a_5)|.\end{aligned}$$

We can now state the following corollary of Theorem 3.2.

THEOREM 4.2. *If conditions* (4.2) *are satisfied, and if*

$$(4.3) \qquad |B_{5,3n-1}| \geqq m\left[\frac{1}{|a_{3n-1}|} + m\,|a_{3n+1}|\right] \qquad (n = 1, 2, 3, \cdots),$$

$$(4.4) \qquad |B_{5,3n}| \geqq m\left[\frac{1}{|a_{3n}|} + m\,|a_{3n+1}|\right] \qquad (n = 1, 2, 3, \cdots),$$

$$(4.5) \qquad |B_{5,3n+1}| \geqq |B_{2,3n+1}| + m^2\,|a_{3n+4}B_{2,3n+4}| \qquad (n = 0, 1, 2, \cdots),$$

the continued fraction (4.1) *converges.*

A simple consequence of this theorem will now be developed. In view of (4.2) we shall suppose henceforth that $a_n \neq -1$ $(n = 2, 3, \cdots)$ and that

$$0 < \delta \leqq |a_{3n+\lambda}| \leqq m \leqq \tfrac{1}{2} \qquad (\lambda = 2, 3;\ n = 0, 1, 2, \cdots), \tag{4.21}$$

where δ and m are independent of n and λ.

Write relations (4.3), (4.4) and (4.5) in expanded form:

$$\begin{aligned} &|a_{3n+2}(1 + a_{3n+4}) + (1 + a_{3n+1})(1 + a_{3n+3} + a_{3n+4})| \\ &\qquad\qquad \geqq m\left[\frac{1}{|a_{3n-1}|} + m\,|a_{3n+1}|\right], \end{aligned} \tag{4.3$'$}$$

$$\begin{aligned} &|a_{3n+3}(1 + a_{3n+5}) + (1 + a_{3n+2})(1 + a_{3n+4} + a_{3n+5})| \\ &\qquad\qquad \geqq m\left[\frac{1}{|a_{3n}|} + m\,|a_{3n+1}|\right], \end{aligned} \tag{4.4$'$}$$

$$\begin{aligned} &|a_{3n+4}(1 + a_{3n+6}) + (1 + a_{3n+3})(1 + a_{3n+5} + a_{3n+6})| \\ &\qquad\qquad \geqq |1 + a_{3n+3}| + m^2\,|a_{3n+4}(1 + a_{3n+6})|. \end{aligned} \tag{4.5$'$}$$

We shall endeavor to replace each of these conditions by simpler and less general conditions which when fulfilled will imply that conditions (4.3)′, (4.4)′ and (4.5)′ respectively are fulfilled. To that end we observe that conditions (4.5)′ will be satisfied if

$$\begin{aligned} &|a_{3n+4}(1 + a_{3n+6})| - |(1 + a_{3n+5} + a_{3n+6})(1 + a_{3n+3})| \\ &\qquad\qquad \geqq |1 + a_{3n+3}| + m^2\,|a_{3n+4}(1 + a_{3n+6})|, \end{aligned}$$

which in turn will be true if

$$|a_{3n+4}|(1 - m) \geqq (1 + m)[1 + m^2\,|a_{3n+4}| + (1 + 2m)]. \tag{4.5$''$}$$

Finally (4.5)″ will be satisfied if the conditions

$$|a_{3n+4}| \geqq \frac{2(1 + m)^2}{1 - m - m^2 - m^3} \qquad (m \leqq \tfrac{1}{2};\ n = 0, 1, 2, \cdots) \tag{4.5$'''$}$$

are satisfied.

We turn our attention to conditions (4.4)′ and observe that (4.4)′ will be satisfied if

$$|a_{3n+4}|(1 - m) \geqq (1 + m)^2 + m(1 + m) + m\left[\frac{1}{\delta} + m\,|a_{3n+1}|\right], \tag{4.4$''$}$$

or if

$$|a_{3n+4}| \geqq \frac{(1 + m)(1 + 2m)}{1 - m} + \frac{t}{1 - m} + m^2\,|a_{3n+1}| \qquad (t \geqq 1), \tag{4.4$'''$}$$

where $m = \delta t$ and $n = 1, 2, 3, \cdots$.

Conditions (4.3)′ will now be treated in somewhat similar fashion. One can verify readily that the following conditions imply (4.3)′:

$$(4.3)'' \quad \begin{aligned} &|a_{3n+4}|\cdot|1 + a_{3n+1} + a_{3n+2}| \\ &\qquad \geqq t + m^2|a_{3n+1}| + |1 + a_{3n+1} + a_{3n+2} + a_{3n+3} + a_{3n+1}a_{3n+3}|. \end{aligned}$$

Further conditions (4.3)″ are implied by conditions

$$(4.3)''' \quad |a_{3n+4}| \geqq 1 + \frac{t + m^2|a_{3n+1}| + m + m^2}{|a_{3n+1}| - 1 - m}$$

$$(m = \delta t,\ t \geqq 1;\ n = 1, 2, 3, \cdots).$$

Conditions (4.5)‴ imply $|a_{3n+4}| > 2(1 + m)$ $(n = 0, 1, 2, \cdots)$. Thus the right member of (4.3)‴ is less than $[m^2|a_{3n+1}| + 1 + 2m + t + m^2]/(1 + m)$ $(n = 1, 2, 3, \cdots)$. It is easy to see that the right member of (4.4)‴ for each n is greater than the right member of (4.3)‴. Thus, if conditions (4.4)‴ are satisfied, conditions (4.3)‴ are automatically fulfilled.

Further, it is easily verified that if $|a_4|$ satisfies (4.5)‴, $|a_{3n+4}|$ also satisfies (4.5)‴ $(n = 1, 2, 3, \cdots)$. We conclude then that the condition

$$(4.51) \qquad |a_4| \geqq \frac{2(1 + m)}{1 - m - m^2 - m^3}$$

and (4.4)‴ together imply (4.3), (4.4), and (4.5).

We proceed to replace the final three "initial conditions" (4.2) by simpler and somewhat less general conditions. The first of these conditions $|B_3| \geqq 1 + |a_1(1 + a_3)|$ can be written

$$(4.6) \qquad |a_1| \leqq \frac{|1 + a_2 + a_3| - 1}{|1 + a_3|}.$$

The second of the last three conditions (4.2) can be written

$$(4.7) \qquad |1 + a_2 + a_3 + a_4 + a_2a_4| \geqq 1 + |a_1a_2(1 + a_4)|.$$

One verifies readily enough that if

$$(4.8) \qquad 1 - m - m|a_1| > 0,$$

(4.7) will be satisfied if

$$(4.7)' \qquad |a_4| \geqq \frac{2 + m|a_1| + 2m}{1 - m - m|a_1|}.$$

Finally, the last condition (4.2) will be satisfied if

$$(4.9) \qquad |a_3(1 + a_5) + (1 + a_2)(1 + a_4 + a_5)| \geqq 1 + m^2|a_1|\frac{1 + m}{1 - m}.$$

Condition (4.9) will be satisfied if

$$(4.9)' \qquad |a_4| \geqq \frac{1}{1 - m}\left[2 + 3m + m^2 + m^2|a_1|\frac{1 + m}{1 - m}\right].$$

We thus have proved the following result.

THEOREM 4.4. *If the elements a_n of* (4.1) *satisfy the conditions*

$$a_2 \neq -\tfrac{1}{2}, \quad or \quad a_3 \neq -\tfrac{1}{2},$$

$$(4.10) \qquad 0 < \delta \leqq |a_{3n+\lambda}| \leqq m \leqq \tfrac{1}{2} \qquad (\lambda = 2, 3; n = 0, 1, 2, \cdots),$$

$$|a_4| \geqq M,$$

where M is the largest of the right members of (4.51), (4.7)$'$, *and* (4.9)$'$, *and if*

$$(4.11) \qquad |a_{3n+4}| \geqq \frac{1 + 3m + 2m^2}{1 - m} + \frac{t}{1 - m} + m^2 |a_{3n+1}|$$

$$(m = \delta t, t \geqq 1, n = 1, 2, 3, \cdots),$$

the continued fraction (4.1) *converges.*

The proof of the theorem can now be given briefly. Choose $|a_1|$ so small that (4.6) and (4.8) are satisfied. Now let M be the largest of the right members of (4.51), (4.7)$'$, and (4.9)$'$. Thus conditions (4.10) with this choice of M imply that conditions (4.2) are satisfied. By the remark following (4.51) conditions (4.11) imply (4.3), (4.4), and (4.5). We observe that the convergence of (4.1) is independent of the choice of $a_1 \neq 0$. The rôle of the conditions applied to $|a_1|$ in the proof of the theorem is thus purely catalytic and the conditions applied to this quantity having served their purpose may now be removed. The proof is complete.

It is clear that similar theorems can be obtained for the case $k = 3$ if r_0 is set equal to 1 and to 2 successively. The results will be analogous to Theorem 4.4 with the subscripts suitably advanced in the conditions of that theorem.

To illustrate the preceding theorem let us set $m = \frac{1}{2}$, $\delta = \frac{1}{4}$, and hence $t = 2$. Conditions (4.11) become

$$(4.12) \qquad |a_{3n+4}| \geqq 10 + \tfrac{1}{4} |a_{3n+1}| \qquad (n = 1, 2, 3, \cdots).$$

The right members of (4.51), (4.7)$'$ and (4.9)$'$ become respectively 24, $(6 + |a_1|)/(1 - |a_1|)$, $\frac{1}{2}(15 + 3|a_1|)$. We have the following result.

Example. If the elements a_n of (4.1) *satisfy the conditions*

$$|a_4| \geqq |24|, \qquad a_2 \quad or \quad a_3 \neq -\tfrac{1}{2},$$

$$\tfrac{1}{4} \leqq |a_{3n+\lambda}| \leqq \tfrac{1}{2} \qquad (\lambda = 2, 3; n = 0, 1, 2, \cdots),$$

$$|a_{3n+4}| \geqq 10 + \tfrac{1}{4} |a_{3n+1}| \qquad (n = 1, 2, 3, \cdots),$$

the continued fraction (4.1) *converges.*

The example provides another set of convergent continued fractions of the form (4.1) all of the elements of which can be greater than $\frac{1}{4}$ in absolute value. It is easily seen that the conditions of Theorems 4.3 and 4.4 are thus independent of earlier known conditions for the convergence of continued fractions (4.1).

Indeed the conditions thus derived from the general theorem by setting $k = 3$ are independent of those conditions (for example (1.1)′) derived from the general theorem by setting $k = 2$.

We conclude with the following result.

THEOREM 4.5. *If the elements a_n of* (4.1) *are any functions of any number of complex variables, the continued fraction* (4.1) *converges uniformly throughout any closed region characterized by the inequalities of Theorem* 3.1. *In particular, if the a_n are analytic functions of a complex variable z, the continued fraction converges to an analytic function of z throughout the interior of any closed region throughout which the inequalities of Theorem* 3.1 *are valid.*

The proof may be given along well-known lines (cf. Perron [1], p. 260) and so is omitted. It is clear that in Theorem 4.5 the conditions of Theorem 3.1 may be replaced by the conditions of any of the theorems of this paper or by those of the example without loss of validity.

BIBLIOGRAPHY

J. Q. JORDAN AND W. LEIGHTON.
1. *On the permutation of the convergents of a continued fraction and related convergence criteria*, Annals of Mathematics, vol. 39(1938), pp. 872–882.

W. LEIGHTON.
1. *Sufficient conditions for the convergence of a continued fraction*, this Journal, vol. 4(1938), pp. 775–778.
2. *A test-ratio test for continued fractions*, Bulletin of the American Mathematical Society, vol. 45(1939), pp. 97–100.

W. LEIGHTON AND H. S. WALL.
1. *On the transformation and convergence of continued fractions*, American Journal of Mathematics, vol. 58(1936), pp. 267–281.

O. PERRON.
1. *Die Lehre von den Kettenbrüchen*, Leipzig, 1929.

O. SZÁSZ.
1. *Über die Erhaltung der Konvergenz Kettenbrüche bei independenter Veränderlichkeit aller ihrer Elemente*, Journal für Mathematik, vol. 147(1916), pp. 132–160.

THE RICE INSTITUTE.

GENERALIZED PROBLEM OF BOLZA IN THE CALCULUS OF VARIATIONS

By M. R. Hestenes

1. **Introduction.** The problem to be studied in the present paper is that of minimizing a function

$$I(C) = g(a) + \int_{x_1}^{x_2} f(a, x, y, y')\,dx \tag{1.1}$$

in a class of admissible arcs C of the form

$$a_h, \quad y_i(x) \qquad (x_1 \leqq x \leqq x_2\,;\, h = 1, \cdots, r;\, i = 1, \cdots, n) \tag{1.2}$$

in axy-space satisfying the conditions

$$\varphi_\gamma(a, x, y, y') = 0 \qquad (\gamma = 1, \cdots, m < n), \tag{1.3a}$$

$$x_s = x_s(a), \quad y_i(x_s) = y_{is}(a) \qquad (s = 1, 2), \tag{1.3b}$$

$$I_\rho = g_\rho(a) + \int_{x_1}^{x_2} f_\rho(a, x, y, y')\,dx = 0 \qquad (\rho = 1, \cdots, p). \tag{1.3c}$$

The a's are independent of the variable x. In the following pages it will be convenient to designate this problem as *problem* A.

The problem just formulated can be modified in many ways. For example, one can suppose that the functions $x_s(a)$ are constants, since this result can be brought about by replacing x by a new variable t by means of the transformation $x = x_1(a) + t[x_2(a) - x_1(a)]$ $(0 \leqq t \leqq 1)$. Moreover, one can assume that the functions $g(a)$, $g_\rho(a)$ are identically zero, for along an admissible arc C satisfying the conditions (1.3) the function (1.1) can be put in the form

$$I(C) = \int_{x_1}^{x_2} \{f + g/[x_2(a) - x_1(a)]\}\,dx$$

and a similar expression holds for $I_\rho(C)$. The simplicity of these transformations of problem A is due to the presence of the a's in the functions $f, f_\rho, \varphi_\gamma$. The introduction of the a's in these functions not only enlarges the class of problems that are immediate special cases of our problem, but also gives rise to a more symmetric theory. This is particularly true in the theory of Mayer fields, as will be seen in §3 below. However, problem A can be reduced to one in which the functions $f, f_\rho, \varphi_\gamma$ are independent of the a's. This can be done by replacing the variables $a_1, \cdots, a_r$ in these functions by new variables $y_{n+1}(x), \cdots, y_{n+r}(x)$ satisfying the conditions $y'_{n+h} = 0$, $y_{n+h}(x_s) = a_h$ $(s = 1, 2)$. If one

Received November 16, 1938.

also replaced the values a_h appearing in the functions g, g_ρ and the end conditions (1.3b) by $y_{n+h}(x_1)$, one would obtain a problem that is a special case of that of minimizing a function

$$I = g[x_1, y(x_1), x_2, y(x_2)] + \int_{x_1}^{x_2} f(x, y, y')\, dx$$

in a class of arcs

$$y_i(x) \qquad (x_1 \leqq x \leqq x_2 ; i = 1, \cdots, n)$$

satisfying conditions of the form

$$\varphi_\gamma(x, y, y') = 0 \qquad (\gamma = 1, \cdots, m < n),$$

$$I_\rho = g_\rho[x_1, y(x_1), x_2, y(x_2)] + \int_{x_1}^{x_2} f_\rho(x, y, y')\, dx = 0.$$

This general problem can also be reduced to one of type A by adjoining the conditions

$$x_1 = a_1, \quad y_i(x_1) = a_{i+1}, \quad x_2 = a_{n+2}, \quad y_i(x_2) = a_{n+i+2}$$

and expressing the functions g, g_ρ in terms of the a's by the use of these equations.

The problem of Bolza as formulated by Bliss (I, II)[1] is the special case of the problem described at the end of the last paragraph in which the functions f_ρ are identically zero. Similarly, the problem of Bolza as formulated by Morse (VIII) is the special case of problem A in which the conditions $I_\rho = 0$ are absent and the functions f, φ_γ are independent of the a's. Conversely, problem A can be reduced to one of Bolza type, namely, to that of minimizing the function (1.1) in the class of admissible arcs

$$a_h, \qquad y_j(x) \qquad (x_1 \leqq x \leqq x_2 ; j = 1, \cdots, n + p)$$

satisfying the conditions (1.3a), (1.3b) and

$$f_\rho(a, x, y, y') - y'_{n+\rho} = 0, \quad y_{n+\rho}(x_1) = g_\rho(a), \quad y_{n+\rho}(x_2) = 0.$$

The a's appearing in the functions f, f_ρ, φ_γ can be replaced by new variables in the manner described above. The concepts of strong relative minima for problem A and the reduced problem of Bolza are, however, not equivalent concepts. It follows that one cannot obtain a complete theory for problem A from that for the problem of Bolza without additional arguments. In order to obtain a strong sufficiency theorem for problem A from that for the problem of Bolza one needs an effective theorem of Lindeberg. Such a theorem has been established recently by W. T. Reid (X; cf. V) who applies it to the problem described at the end of the second paragraph of this paper. A similar difficulty is encountered when one applies the theory of the problem of Bolza to the isoperimetric problem and to the problem of Euler recently studied by Brady (III).

[1] Roman numerals in parentheses refer to the references at the end of this paper.

On the other hand the problem of Bolza and the isoperimetric problems are immediate special cases of problem A. Similarly, the problem of Euler studied by Brady is the special case of problem A in which $f \equiv 0$, $g_\rho = -a_\rho$, the functions f_ρ, x_s, y_{is} are independent of the a's and the differential equations $\varphi_\gamma = 0$ are absent. For these problems and others of similar type one can obtain complete sets of necessary conditions and sufficient conditions for strong relative minima from those for problem A without further arguments.

During the last few years the problem of Bolza has been studied in great detail and the results obtained have been applied to problems that can be reduced to one of Bolza type. In view of the above remarks it appears to the author that it would be more economical and satisfactory to study first the generalized problem of Bolza here formulated or one of similar type and then apply the results so obtained to the various special cases. Most of the theory for problem A can be established by simple modifications of the arguments used to develop the corresponding theory for the problem of Bolza. A new sufficiency proof is needed, however. It is the purpose of the present paper to give such a proof. The method used is an extension of one recently used by the author (VI) for the problem of Bolza. Numerous simplifications of the earlier method have been made.

2. **Preliminary remarks.** In the following pages it will be assumed that the functions appearing in the expressions (1.1) and (1.3) are continuous and have continuous partial derivatives of the first three orders in a region $\mathfrak{R}$ of points (a, x, y, y'). We suppose further that $x_1(a) < x_2(a)$ and the matrix $\| \varphi_{\gamma y_i'} \|$ has rank m on $\mathfrak{R}$. The elements (a, x, y, y') in $\mathfrak{R}$ will be called *admissible*. By an *admissible arc* C will be meant a continuous arc (1.2) in axy-space that can be subdivided into a finite number of subarcs on each of which it has continuous derivatives and has its elements (a, x, y, y') admissible. This definition of admissible arcs is not the one previously used by the author, but is essentially the one used by Bliss (I, p. 10).

By an *extremal* E will be meant an admissible arc (1.2) without corners and a set of multipliers l_ρ, $m_\gamma(x)$ having continuous derivatives y_i'', $l_\rho' = 0$, m_γ' and satisfying the Euler-Lagrange equations

$$F_{y_i} - \frac{d}{dx} F_{y_i'} = 0, \qquad \varphi_\gamma = 0 \qquad (\gamma = 1, \cdots, m), \tag{2.1}$$

where

$$F(a, x, y, y', l, m) = f + l_\rho f_\rho + m_\gamma \varphi_\gamma .$$

An extremal E will be said to be *non-singular* if the determinant

$$\begin{vmatrix} F_{y_i' y_k'} & \varphi_{\gamma y_i'} \\ \varphi_{\beta y_k'} & 0 \end{vmatrix} \tag{2.2}$$

is different from zero along E. From existence theorems for differential equations applied to equations (2.1) one obtains the following result (cf. I, pp. 33–36).

THEOREM 2.1. *Every non-singular extremal E is a member of an $(r + 2n + p)$-parameter family of extremals*

$$a_h, \quad y_i(x, a, b, c, l), \quad l_\rho, \quad m_\gamma(x, a, b, c, l) \qquad (x_1 \leqq x \leqq x_2) \tag{2.3}$$

for special values $a_h = a_{h0}$, $b_i = b_{i0}$, $c_i = c_{i0}$, $l_\rho = l_{\rho 0}$, $x_{10} \leqq x \leqq x_{20}$, where $h = 1, \cdots, r$; $i = 1, \cdots, n$; $\rho = 1, \cdots, p$. The functions y_i, y_{ix}, m_γ, $z_i = F_{y'_i}$, z_{ix} of (x, a, b, c, l) determined by this family are continuous and have continuous first and second derivatives with respect to their arguments in a neighborhood of the values (x, a, b, c, l) belonging to E. The determinant

$$\begin{vmatrix} y_{ib_j} & y_{ic_j} \\ z_{ib_j} & z_{ic_j} \end{vmatrix} \qquad (i, j = 1, \cdots, n)$$

is different from zero along E. The parameters b_i, c_i can be chosen to be the values of y_i, z_i at a fixed value of x.

The principal theorem to be proved in the present paper is the following one. Precise definitions of the terms used will be given below.

THEOREM 2.2. SUFFICIENT CONDITIONS FOR A STRONG RELATIVE MINIMUM. *If a non-singular extremal E satisfies the conditions* (1.3), *the transversality condition* (2.4), *the Weierstrass condition* II_N, *and is such that the second variation of I along E is positive, then there is a neighborhood $\mathfrak{F}$ of E in axy-space such that the inequality $I(C) > I(E)$ holds for every admissible arc C in $\mathfrak{F}$ satisfying the conditions* (1.3) *and not identical with E.*

An extremal E is said to satisfy the *transversality condition* if every set of constants $da_1, \cdots, da_r$ satisfies with E the equation

$$dg + l_\rho dg_\rho + [(F - y'_i F_{y'_i})\,dx + F_{y'_i}\,dy_i]_1^2 + \int_{x_1}^{x_2} F_{a_h}\,da_h\,dx = 0, \tag{2.4}$$

where dx_1, dy_{i1}, dx_2, dy_{i2} are the differentials of the second members of equations (1.3b).

An extremal E is said to satisfy the *Weierstrass condition* II_N if at each element (a, x, y, y', l, m) in a neighborhood N of those on E and satisfying the conditions $\varphi_\gamma = 0$ the inequality

$$E(a, x, y, y', l, m, Y') \geqq 0 \tag{2.5}$$

holds for every set (Y') such that the element (a, x, y, Y') is admissible and satisfies the conditions $\varphi_\gamma = 0$. Here

$$E = F(Y') - F(y') - (Y'_i - y'_i)F_{y'_i}(y'), \tag{2.6}$$

where the arguments in F and its derivatives not indicated are (a, x, y, l, m). For a non-singular arc E satisfying the condition II_N the equality in (2.5) will

hold only in case $(Y') = (y')$ if one chooses the neighborhood N so that the determinant (2.2) is different from zero on N. This result has been established recently by Reid and the author (VII).

By the second variation of J along an extremal will be meant the expression (cf. VIII, pp. 520–521)

$$J(\alpha, \eta) = b_{hk}\alpha_h\alpha_k + \int_{x_1}^{x_2} 2\omega(\alpha, x, \eta, \eta')\,dx, \tag{2.7}$$

where $h, k = 1, \cdots, r$; $i, j = 1, \cdots, n$ and

$$\begin{aligned} b_{hk} = {} & g_{hk} + l_\rho g_{\rho hk} + [(F - y_i'F_{y_i'})x_{shk} + F_{y_i'}y_{ishk}]_{s=1}^{s=2} \\ & + [(F_x - y_i'F_{y_i})x_{sh}x_{sk} + F_{y_i}(x_{sh}y_{isk} + x_{sk}y_{ish}) + x_{sh}F_{a_k} + x_{sk}F_{a_h}]_{s=1}^{s=2}, \end{aligned} \tag{2.8a}$$

$$\begin{aligned} 2\omega = {} & F_{y_iy_j}\eta_i\eta_j + 2F_{y_iy_j'}\eta_i\eta_j' + F_{y_i'y_j'}\eta_i'\eta_j' + 2F_{y_ia_h}\eta_i\alpha_h \\ & + 2F_{y_i'a_h}\eta_i'\alpha_h + F_{a_ha_k}\alpha_h\alpha_k . \end{aligned} \tag{2.8b}$$

Here and elsewhere the subscripts h, k on g, g_ρ, x_s, y_{is} denote derivatives of the functions $g(a)$, $g_\rho(a)$, $x_s(a)$, $y_{is}(a)$ with respect to a_h, a_k at the values of (a) belonging to E. It is understood that the constants α_h and the functions $\eta_i(x)$ define an arc in $\alpha x\eta$-space with continuity properties like those of admissible arcs. These arcs will be called *admissible variations*. The second variation J of I will be said to be positive along E if the inequality $J(\alpha, \eta) > 0$ holds for every set of admissible variations $(\alpha, \eta) \neq (0, 0)$ satisfying with E the conditions

$$\Phi_\gamma(\alpha, x, \eta, \eta') = \varphi_{\gamma a_h}\alpha_h + \varphi_{\gamma y_i}\eta_i + \varphi_{\gamma y_i'}\eta_i' = 0 \qquad (\gamma = 1, \cdots, m), \tag{2.9a}$$

$$\eta_i(x_s) = \eta_{ish}\alpha_h \qquad (s = 1, 2), \tag{2.9b}$$

$$J_\rho(\alpha, \eta) = c_{\rho h}\alpha_h + \int_{x_1}^{x_2} \omega_\rho(\alpha, x, \eta, \eta')\,dx = 0 \; (\rho = 1, \cdots, p), \tag{2.9c}$$

where

$$\eta_{ish} = y_{ish} - y_i'(x_s)x_{sh} \qquad (s = 1, 2;\ s \text{ not summed}), \tag{2.10}$$

$$c_{\rho h} = g_{\rho h} + [f_\rho x_{sh}]_{s=1}^{s=2}, \qquad \omega_\rho = f_{\rho a_h}\alpha_h + f_{\rho y_i}\eta_i + f_{\rho y_i'}\eta_i' . \tag{2.11}$$

By an *accessory extremal* will be meant an admissible variation and a set of multipliers

$$\alpha_h, \quad \eta_i(x), \quad \lambda_\rho, \quad \mu_\gamma(x) \qquad (x_1 \leqq x \leqq x_2) \tag{2.12}$$

having continuous derivatives η_i', η_i'', $\lambda_\rho' = 0$, μ_γ' and satisfying the *accessory differential equations*

$$\Omega_{\eta_i} - \frac{d}{dx}\Omega_{\eta_i'} = 0, \qquad \Phi_\gamma = 0, \tag{2.13}$$

where

$$\Omega = \omega + \lambda_\rho\omega_\rho + \mu_\gamma\Phi_\gamma \tag{2.14}$$

and the functions $\omega, \omega_\rho, \Phi_\gamma$ are defined by equations (2.8b), (2.11), (2.9a), respectively. If E is non-singular, an accessory extremal is uniquely determined by the values of the functions $\alpha_h, \eta_i, \zeta_i = \Omega_{\eta_i'}, \lambda_\rho$ at a point $x = x_0$. We may denote therefore an accessory extremal by the symbols $\alpha_h, \eta_i, \zeta_i, \lambda_\rho$.

A set of functions $u_{ij}(x), v_{ij}(x)$ $(j = 1, \cdots, n)$ will be said to form a *conjugate system* for a non-singular extremal E if there exists a set of n linearly independent accessory extremals of the form

$$\alpha_{hj} = 0, \quad \eta_{ij} = u_{ij}, \quad \zeta_{ij} = v_{ij}, \quad \lambda_{\rho j} = 0 \qquad (j = 1, \cdots, n) \tag{2.15}$$

such that $u_{ij}v_{ik} - v_{ij}u_{ik} = 0$ $(i, j, k = 1, \cdots, n)$. It is well known that if the last equation holds at one value of x, it holds for all values of x on x_1x_2 (I, p. 80).

The following lemma will be useful in §5 below.

LEMMA 2.1. *If E is a non-singular extremal, there is a constant $\delta > 0$ such that for every subinterval $x'x''$ on x_1x_2 of length at most δ there exists a conjugate system u_{ij}, v_{ij} for E having $|u_{ij}(x)| \neq 0$ on $x'x''$.*

For, if x_0 is a value on x_1x_2, then a set of accessory extremals of the form (2.15) with $|\eta_{ij}(x_0)| \neq 0$ defines a conjugate system u_{ij}, v_{ij} with $|u_{ij}(x)| \neq 0$ on a neighborhood of $x = x_0$. The lemma now follows by an application of the Heine-Borel Theorem.

3. **Mayer fields.** A region $\mathfrak{F}$ in axy-space and a set of slope functions $p_i(a, x, y)$ and multipliers $l_\rho(a)$, $m_\gamma(a, x, y)$, that are continuous and have continuous derivatives of the first two orders, will be said to define a *Mayer field* $\mathfrak{F}$ if the sets (a, x, y, p) are admissible and satisfy the equations $\varphi_\gamma(a, x, y, p) = 0$ and the expression

$$I^*(C) = g^*(a) + \int_{x_1}^{x_2} F^*(a, x, y, y')\,dx, \tag{3.1}$$

where $g^* = g + l_\rho g_\rho$ and

$$F^* = F(a, x, y, p, l, m) + (y_i' - p_i)F_{y_i'}(a, x, y, p, l, m)$$

is independent of the path in $\mathfrak{F}$ in the sense that the value $I^*(C)$ is the same for all admissible arcs C in $\mathfrak{F}$ having the same end values $[a, x_1, y(x_1), x_2, y(x_2)]$. For an admissible arc C in $\mathfrak{F}$ satisfying the conditions (1.3c) one has the formula

$$I(C) = I^*(C) + \int_C E(a, x, y, p, l, m, y')\,dx, \tag{3.2}$$

where E is the Weierstrass E-function (2.6).

A solution (1.2) of the equations $y_i' = p_i(a, x, y)$ can be shown (cf. I, pp. 102-103) to form an extremal with the multipliers $l_\rho(a)$, $m_\gamma[a, x, y(x)]$. Such an extremal will be called an *extremal of the field*. Through each point (a, x, y) in $\mathfrak{F}$ there passes one and only one extremal of the field. Moreover, from the formula (3.2) it follows that the relation $I(E) = I^*(E)$ holds for every extremal of the field satisfying the conditions (1.3c).

THEOREM 3.1. *Let E be an extremal of a Mayer field $\mathfrak{F}$ at each point of which the inequality*

$$E[a, x, y, p(a, x, y), l(a), m(a, x, y), y'] > 0 \tag{3.3}$$

holds for every set $(y') \neq (p)$ such that (a, x, y, y') is admissible and satisfies the equations $\varphi_\gamma = 0$. Suppose further that the relation $I^*(C) \geqq I^*(E)$ holds for every admissible arc C in $\mathfrak{F}$ satisfying the end conditions (1.3b), the equality holding only in case C and E have the same components a_h. Then the inequality $I(C) > I(E)$ holds for every admissible arc C in $\mathfrak{F}$ satisfying the conditions (1.3) and not identical with E.

For by virtue of the formula (3.2) and the hypotheses of the theorem one has $I(C) \geqq I^*(C) \geqq I^*(E) = I(E)$ for every admissible arc C in $\mathfrak{F}$ satisfying the conditions (1.3). The equality holds throughout only in case $y_i' = p_i$ along C and the arcs C and E have the same components a_h and hence the same initial point, by equations (1.3b). But this implies that C is an extremal of the field with the same initial point as E. The arc C is therefore identical with E and the theorem is proved.

The above theorem suggests the study of the problem of minimizing $I^*(C)$ in the class of admissible arcs C satisfying the conditions (1.3b) but not necessarily the conditions (1.3a) and (1.3c). Suppose now that E is an extremal of a Mayer field satisfying the conditions (1.3) and minimizing I^* subject to the conditions (1.3b). Then E must satisfy the necessary conditions for a minimum for this problem. It is clear that E is an extremal relative to I^*. The transversality condition for this new problem is obtained by replacing F, g in (2.4) by F^*, g^*, respectively. By the use of equations (1.3) and $y_i' = p_i$ it is found that E satisfies this new transversality condition if and only if it satisfies the condition (2.4). Moreover, the second variation $J^*(\alpha, \eta)$ of I^* along E must be non-negative for every admissible variation (α, η) satisfying the conditions (2.9b). By a somewhat complicated but not difficult computation it is found that along E

$$J^*(\alpha, \eta) = b_{hk}\alpha_h\alpha_k + 2\lambda_\rho c_{\rho h}\alpha_h + 2\int_{x_1}^{x_2} \{\Omega + (\eta_i' - \pi_i)\Omega_{\eta_i'}\}\,dx, \tag{3.4}$$

where b_{hk}, $c_{\rho h}$, Ω are defined by equations (2.8a), (2.11), (2.14) and the arguments in Ω and $\Omega_{\eta_i'}$ are $(\alpha, x, \eta, \pi, \lambda, \mu)$, the values π_i, λ_ρ, μ_γ being determined by the equations

$$\pi_i = p_{ia_h}\alpha_h + p_{iy_j}\eta_j, \quad \lambda_\rho = l_{\rho a_h}\alpha_h, \quad \mu_\gamma = m_{\gamma a_h}\alpha_h + m_{\gamma y_j}\eta_j. \tag{3.5}$$

The second variation J^* of I^* along E is also expressible in the form

$$J^*(\alpha, \eta) = b_{hk}\alpha_h\alpha_k + 2\lambda_\rho c_{\rho h}\alpha_h + \int_{x_1}^{x_2} \{2\Omega - F_{y_i'y_k'}(\eta_i' - \pi_i)(\eta_k' - \pi_k)\}\,dx, \tag{3.6}$$

where the arguments in Ω are now $(\alpha, x, \eta, \eta', \lambda, \mu)$. The equivalence of the formulas (3.4) and (3.6) is easily established by expanding $2\Omega[\alpha, x, \eta, \pi + (\eta' - \pi), \lambda, \mu]$ in terms of $\eta_i' - \pi_i$ by means of Taylor's formula.

THEOREM 3.2. *Let E be an extremal of a Mayer field satisfying the conditions* (1.3) *and* (2.4). *Suppose that the second variation $J^*(\alpha, \eta)$ of I^* along E satisfies the condition $J^*(\alpha, \eta) > 0$ for every admissible arc $(\alpha, \eta) \neq (0, \eta)$ satisfying the conditions* (2.9b). *Then there is a neighborhood $\mathfrak{F}_1$ of E in axy-space such that the inequality $I^*(C) \geqq I^*(E)$ holds for every admissible arc C in $\mathfrak{F}$ satisfying the conditions* (1.3b), *the equality holding only in case C and E have the same components a_h.*

To prove this result we may suppose that the components a_h belonging to E are given by the set $a_h = 0$. Let α_{hk}, η_{ik} $(k = 1, \cdots, r)$ be a set of r admissible variations having continuous second derivatives, satisfying the conditions (2.9b) and having $\alpha_{hh} = 1$, $\alpha_{hk} = 0$ $(h \neq k)$. Let

$$y_i(x, a) = Y_i(x, a) + h_{i1}(a)[x_2(a) - x] + h_{i2}(a)[x - x_1(a)],$$

where Y_i, h_{is} are defined by the equations

$$Y_i = y_i(x) + \eta_{ik}a_k, \qquad h_{is}[x_2(a) - x_1(a)] = y_{is}(a) - Y_i[x_s(a), a],$$

the functions $y_i(x)$ being those belonging to E and the functions $x_s(a)$, $y_{is}(a)$ being those appearing in equations (1.3b). The r-parameter family of admissible arcs

$$a_h, \qquad y_i(x, a) \qquad (x_1(a) \leqq x \leqq x_2(a)) \tag{3.7}$$

so obtained contains E for values $a_h = 0$, satisfies the end conditions (1.3b) and has α_{hk}, η_{ik} as its variations along E. When the functions (3.7) are substituted in the expression (3.1) for I^*, a function $I^*(a)$ is obtained having continuous first and second derivatives. The value of dI^* at $a_h = 0$ is equal to the value of the first variation of I^* along E determined by the admissible variation $\alpha_h = da_h$, $\eta_i = \eta_{ik}da_k$ and is equal to zero since E is an extremal for I^* satisfying the transversality condition for I^*, as was seen above. Similarly, the value of d^2I^* at $a_h = 0$ is equal to the value of $J^*(\alpha, \eta)$ determined by the admissible variation just described. Hence $d^2I^* > 0$ for all $(da) \neq (0)$ by virtue of our hypothesis concerning $J^*(\alpha, \eta)$. It follows that $I^*(a) > I^*(0) = I^*(E)$ for every set $(a) \neq (0)$ in a neighborhood A of $(a) = (0)$. Let $\mathfrak{F}_1$ be all points (a, x, y) in $\mathfrak{F}$ with (a) in A and consider an admissible arc C in $\mathfrak{F}_1$ satisfying the conditions (1.3b). Since the components a_h belonging to C determine an arc (3.7) joining the ends of C, one has $I^*(C) = I^*(a)$. Hence $I^*(C) \geqq I^*(E)$, the equality holding only in case C and E have the same components a_h. This proves Theorem 3.2.

THEOREM 3.3. *Let E be an extremal of a Mayer field $\mathfrak{F}$ satisfying the conditions* (1.3). *The set of all points (α, x, η) with x on x_1x_2 together with the slope functions π_i and multipliers λ_ρ, μ_γ given by* (3.5) *define an accessory Mayer field for the*

problem of minimizing the second variation $J(\alpha, \eta)$ *of* I *along* E *subject to the conditions* (2.9). *The invariant integral for this field is given by* (3.4).

In order to establish this result let α_h, $\eta_i(x)$ and α_h, $\bar{\eta}_i(x)$ be admissible variations having the same α's and $\eta_i(x_s) = \bar{\eta}_i(x_s)$ $(s = 1, 2)$. Let $a_h = 0$, $y_i(x)$ $(x_1 \leqq x \leqq x_2)$ be the functions defining E and consider the family of admissible arcs

$$a_h(e) = e\alpha_h, \quad y_i(x, e) = y_i(x) + e\eta_i(x) \qquad (x_1 \leqq x \leqq x_2). \tag{3.8}$$

Let $I^*(e)$ be the values of I^* determined by this family, and denote by $\bar{I}^*(e)$ the values of I^* determined by the family obtained from (3.8) when η_i is replaced by $\bar{\eta}_i$. Since the arcs belonging to these two families for a particular value of e have the same end values, the function $H(e) = I^*(e) - \bar{I}^*(e)$ is identically zero. But $H''(0) = J^*(\alpha, \eta) - J^*(\alpha, \bar{\eta})$, as one readily verifies. We have accordingly $J^*(\alpha, \eta) = J^*(\alpha, \bar{\eta})$, as was to be proved.

THEOREM 3.4. *Let E be a non-singular extremal satisfying the conditions* (1.3). *Suppose that there exists an accessory Mayer field for the second variation* $J(\alpha, \eta)$ *of* I *along* E *of the type described in the last theorem. Then* E *is an extremal of a Mayer field* $\mathfrak{F}$ *such that the slope functions and multipliers of the accessory Mayer field are the variations* (3.5) *along* E *of the slope functions and multipliers of* $\mathfrak{F}$.

This theorem can be established by an argument similar to those made in the next section. However, the proof of this theorem will be omitted in view of the fact that we shall make no explicit use of this result.

4. **Construction of fields.** The following theorem establishes the existence of fields of the type described above.

THEOREM 4.1. *Let E be a non-singular extremal for which there exists a conjugate system* u_{ij}, v_{ij} *having* $| u_{ij}(x) | \neq 0$ *along it and let*

$$\alpha_{hk}, \quad \eta_{ik}(x), \quad \lambda_{\rho k}, \quad \mu_{\gamma k}(x) \quad [\alpha_{hh} = 1, \alpha_{hk} = 0 \ (h \neq k); h, k = 1, \cdots, r] \tag{4.1}$$

be a set of r *accessory extremals for* E. *There exists an* $(r + n)$*-parameter family of extremals*

$$a_h, \quad y_i(x, a, e), \quad l_\rho(a), \quad m_\gamma(x, a, e) \tag{4.2}$$

containing E *for values* $x_1 \leqq x \leqq x_2$, $a_h = a_{h0}$ $(h = 1, \cdots, r)$, $e_i = e_{i0}$ $(i = 1, \cdots, n)$. *The functions* y_i, y_{ix}, l_ρ, m_γ, $z_i = F_{y_i'}$ *of* (x, a, e) *determined by this family have continuous first and second derivatives in a neighborhood of the values* (x, a, e) *belonging to* E. *Moreover, along* E *one has*

$$\begin{aligned} y_{ia_k} &= \eta_{ik}, \quad l_{\rho a_k} = \lambda_{\rho k}, \quad m_{\gamma a_k} = \mu_{\gamma k}, \\ y_{ie_j} &= u_{ij}, \quad z_{ie_j} = v_{ij}. \end{aligned} \tag{4.3}$$

The extremal E *is an extremal of a Mayer field* $\mathfrak{F}$ *with slope functions and multipliers*

$$p_i(a, x, y) = y_{ix}[x, a, e(a, x, y)], \quad l_\rho(a), \quad m_\gamma(a, x, y) = m_\gamma[x, a, e(a, x, y)], \tag{4.4}$$

where $e_j(a, x, y)$ is the value of e_j belonging to the extremal (4.2) *passing through the point (a, x, y) in $\mathfrak{F}$.*

To prove this suppose that $a_{h0} = 0$ and that the parameters b_i, c_i in the family (2.3) have been chosen to be the values of y_i, z_i at a point $x = x_0$ on x_1x_2. A family (4.2) having the properties described in the theorem can be obtained from the family (2.3) by setting

$$(4.5) \qquad \begin{aligned} b_i &= b_{i0} + \eta_{ih}(x_0)a_h + u_{ij}(x_0)e_j, \\ c_i &= c_{i0} + \zeta_{ih}(x_0)a_h + v_{ij}(x_0)e_j, \qquad l_\rho = l_{\rho 0} + \lambda_{\rho k}a_k, \end{aligned}$$

where ζ_{ik} are the values of $\Omega_{\eta_i'}$ determined by the accessory extremals (4.1). The continuity properties of the family are immediate. From equations (4.5) and the identities

$$b_i = y_i(x_0, a, e), \qquad c_i = z_i(x_0, a, e)$$

in a_h, e_j it is found by differentiation that equations (4.3) hold at $x = x_0$ and hence along E since these functions are related in a unique manner with accessory extremals. It follows that the determinant $|y_{ie_j}|$ is different from zero along E. The equations $y_i = y_i(x, a, e)$ accordingly have unique solutions $e_j(a, x, y)$ in a neighborhood $\mathfrak{F}$ of E in axy-space. On the hyperspace $x = x_0$, $a_h = \text{const.}$, the Hilbert integral

$$\int \{F(a, x, y, p, l, m)\, dx + (dy_i - p_i\, dx)F_{y_i'}(a, x, y, p, l, m)\}$$

determined by the functions (4.4) takes the form $\int c_i\, db_i = \int dW$, where

$$2W(e) = 2(c_{i0} + \zeta_{ih}a_h)u_{ij}e_j + u_{ij}v_{ik}e_je_k \qquad (i, j, k = 1, \cdots, n)$$

and hence is independent of the path in $\mathfrak{F}$ (I, p. 106). It follows that the region $\mathfrak{F}$ and the functions (4.4) define a Mayer field whose extremals are given by the family (4.2). This proves Theorem 4.1.

THEOREM 4.2. *Suppose the hypotheses of Theorem 4.1 hold and let*

$$(4.6) \qquad \pi_i = \eta_{ik}'\alpha_k + u_{ij}'\epsilon_j, \qquad \lambda_\rho = \lambda_{\rho k}\alpha_k, \qquad \mu_\gamma = \mu_{\gamma k}\alpha_k + \nu_{\gamma j}\epsilon_j,$$

where $\epsilon_j = \epsilon_j(\alpha, x, \eta)$ are the solutions of the equations $\eta_i = \eta_{ik}\alpha_k + u_{ij}\epsilon_j$ and $\nu_{\gamma j}(x)$ are the multipliers μ_γ belonging to the accessory extremals (2.15) *defining the conjugate system u_{ij}, v_{ij}. The functions* (4.6) *are the slope functions and multipliers of the accessory Mayer field associated with the field $\mathfrak{F}$ described in Theorem* 4.1. *In fact if the second variation $J(\alpha, \eta)$ of I along E is positive along E, the accessory extremals* (4.1) *can be chosen so that the invariant integral* (3.4) *for the accessory Mayer field satisfies the condition $J^*(\alpha, \eta) > 0$ for every admissible variation $(\alpha, \eta) \neq (0, \eta)$ satisfying the conditions* (2.9b).

To prove this we note that if in the identity $y_i = y_i[x, a, e(a, x, y)]$ the variables y_i, a_h are replaced by $y_i + b\eta_i$, $a_h + b\alpha_h$ and the result is differentiated for b, one obtains upon setting $b = 0$ and using the relations (4.3) the identity

$$\eta_i = \eta_{ik}\alpha_k + u_{ij}(e_{ja_h}\alpha_h + e_{jy_k}\eta_k)$$

along E. It follows that $\epsilon_j(\alpha, x, \eta)$ are the variations of $e_j(a, x, y)$ along E. The variations of the functions (4.4) along E are therefore given by the set (4.6). The first part of the theorem now follows from Theorem 3.3. The proof of the last statement of the theorem is based on two lemmas, the first of which is the following

LEMMA 4.1. *Let E be a non-singular extremal and*

$$\alpha_{hj}, \quad \eta_{ij}(x), \quad \lambda_{\rho j}, \quad \mu_{\gamma j}(x) \qquad (j = 1, \cdots, t) \tag{4.7}$$

a maximal set of accessory extremals for E such that the variations α_{hj}, η_{ij} are linearly independent on x_1x_2. If the second variation $J(\alpha, \eta)$ of I is positive along E, the matrix whose j-th row is given by the set

$$\alpha_{hj}, \quad \eta_{ij}(x_2), \quad \eta_{ij}(x_1), \quad J_\rho(\alpha_j, \eta_j) \tag{4.8}$$

has rank t. Moreover $t = r + 2n + p - q$, where q is the number of linearly independent accessory extremals α_h, η_i, λ_ρ, μ_γ in a maximal set having $\alpha_h = \eta_i(x) \equiv 0$ on x_1x_2.

For suppose the matrix described in the lemma did not have rank t. Then there would exist multipliers β_j, not all zero, such that

$$\alpha_{hj}\beta_j = \eta_{ij}(x_2)\beta_j = \eta_{ij}(x_1)\beta_j = J_\rho(\alpha_j, \eta_j)\beta_j = 0.$$

The accessory extremal

$$\alpha_h = \alpha_{hj}\beta_j, \qquad \eta_i = \eta_{ij}\beta_j, \qquad \lambda_\rho = \lambda_{\rho j}\beta_j, \qquad \mu_\gamma = \mu_{\gamma j}\beta_j$$

would have $\alpha_h = \eta_i(x_1) = \eta_i(x_2) = J_\rho(\alpha, \eta) = 0$ and $\eta \not\equiv 0$ on x_1x_2. By an integration by parts with the help of equations (2.13) it would be found that

$$J(\alpha, \eta) = J(\alpha, \eta) + 2\lambda_\rho J_\rho(\alpha, \eta) = \int_{x_1}^{x_2} 2\Omega\, dx = [\eta_i \Omega_{\eta_i'}]_1^2 = 0,$$

contrary to our assumption that the second variation is positive along E. This proves the first part of the lemma. The last statement follows from the fact that there are $r + 2n + p$ linearly independent accessory extremals in a maximal set.

Consider now the expression

$$H_\sigma(\alpha, \eta) = l_{\rho\sigma} J_\rho(\alpha, \eta) + \int_{x_1}^{x_2} m_{\gamma\sigma}\Phi_\gamma(\alpha, \eta, \eta')\, dx \qquad (\sigma = 1, \cdots, q), \tag{4.9}$$

where $l_{\rho\sigma}$, $m_{\gamma\sigma}$ are the multipliers belonging to a maximal set of linearly independent accessory extremals α_h, η_i, λ_ρ, μ_γ of the form

$$(4.10)\qquad a_{h\sigma} = 0, \qquad y_{i\sigma} = 0, \qquad l_{\rho\sigma}, \qquad m_{\gamma\sigma} \qquad (\sigma = 1, \cdots, q).$$

Let $z_{i\sigma}(x)$ be the corresponding values of $\Omega_{\eta_i'}$ and let $w_{\sigma h}$ be the coefficient of α_h in the expression (4.9). By the use of the accessory equations (2.13) for the accessory extremals (4.10) it is found that

$$(4.11)\qquad H_\sigma(\alpha, \eta) = w_{\sigma h}\alpha_h + \int_{x_1}^{x_2} \{z_{i\sigma}'\eta_i + z_{i\sigma}\eta_i'\}\, dx = w_{\sigma h}\alpha_h + [z_{i\sigma}\eta_i]_1^2.$$

The second lemma to be established is the following

Lemma 4.2. *If $\bar{\alpha}_h$, $\bar{\eta}_i$ is an admissible variation satisfying with a set of constants J_ρ the equations $H_\sigma(\alpha, \eta) = l_{\rho\sigma}J_\rho$, there exists an accessory extremal α_h, η_i, λ_ρ, μ_γ such that*

$$(4.12)\qquad \alpha_h = \bar{\alpha}_h, \qquad \eta_i(x_s) = \bar{\eta}_i(x_s)\ (s = 1, 2), \qquad J_\rho(\alpha, \eta) = J_\rho.$$

In particular if $\Phi_\gamma(\bar{\alpha}, x, \bar{\eta}, \bar{\eta}') = 0$, then $J_\rho(\alpha, \eta) = J_\rho(\bar{\alpha}, \bar{\eta})$.

For since the accessory extremals (4.10) are linearly independent, the matrix $\| z_{i\sigma}(x) \quad l_{\rho\sigma} \|$ has rank q. The equations

$$w_{\sigma h}\alpha_h + z_{i\sigma}(x_2)\eta_{i2} - z_{i\sigma}(x_1)\eta_{i1} = l_{\rho\sigma}J_\rho$$

are therefore linearly independent equations in the variables α_h, η_{i2}, η_{i1}, J_ρ and have $t = r + 2n + p - q$ linearly independent solutions. By virtue of the relations (4.9) and (4.11) a maximal set of linearly independent solutions of these equations is given by the set (4.8). There is accordingly a linear combination α_h, η_i, λ_ρ, μ_γ of the extremals (4.7) satisfying equations (4.12). This proves the lemma.

We are now in position to complete the proof of Theorem 4.2. To do so let

$$(4.13)\qquad \alpha_{hk}, \qquad \eta_{ik}(x), \qquad \lambda_{\rho k}, \qquad \mu_{\gamma k}(x) \qquad (k = 1, \cdots, r)$$

be a set of r accessory extremals having $|\alpha_{hk}| \neq 0$ and such that the last $r - r'$ of these form a maximal set of accessory extremals satisfying the conditions (2.9) and having its matrix $\| \alpha_{hl} \|$ $(l = r' + 1, \cdots, r)$ of rank $r - r'$. We may suppose without loss of generality that $\alpha_{hh} = 1$, $\alpha_{hk} = 0$ $(h \neq k)$ since this choice is always possible if we first transform the a's by the transformation $a_h = a_{h0} + \alpha_{hk}\bar{a}_k$, where a_{h0} are the values of a_h belonging to E. Let $\bar{\alpha}_{hk}$, $\bar{\eta}_{ik}$ $(k = 1, \cdots, r)$ be a set of r admissible variations satisfying the conditions (2.9b) and having $\bar{\alpha}_{hk} = \alpha_{hk}$, $\bar{\eta}_{il} = \eta_{il}$ $(l = r' + 1, \cdots, r)$. It is clear from equations (4.9) and (2.9) that the values $H_{\sigma k} = H_\sigma(\bar{\alpha}_k, \bar{\eta}_k)$ are zero when $k > r'$. However, the matrix $\| H_{\sigma k} \|$ has rank r'. Otherwise there would exist constants β_τ $(\tau = 1, \cdots, r')$, not all zero, such that $H_{\sigma\tau}\beta_\tau = H_\sigma(\bar{\alpha}_\tau\beta_\tau, \bar{\eta}_\tau\beta_\tau) = 0$. But by Lemma 4.2 with $J_\rho = 0$ there would exist an accessory extremal α_h, η_i, λ_ρ, μ_γ satisfying the equations

$$\alpha_h = \bar{\alpha}_{h\tau}\beta_\tau, \qquad \eta_i(x_s) = \bar{\eta}_{i\tau}(x_s)\beta_\tau\ (s = 1, 2), \qquad J_\rho(\alpha, \eta) = 0$$

and hence also equations (2.9), contrary to our choice of the last $r - r'$ of the accessory extremals (4.13).

Consider now the accessory extremals

$$(4.14) \qquad \alpha_{hk}, \qquad \eta_{ik}(x), \qquad \lambda_{\rho k} + bl_{\rho\sigma}H_{\sigma k}, \qquad \mu_{\gamma k} + bm_{\gamma\sigma}H_{\sigma k},$$

where $l_{\rho\sigma}$, $m_{\gamma\sigma}$ are the multipliers belonging to the extremals (4.10) and b is a constant to be chosen below. Let $J^*(\alpha, \eta)$ and $J_1^*(\alpha, \eta)$ be the invariant integrals (3.6) for the accessory Mayer fields determined respectively by the accessory extremals (4.13) and (4.14) as described in the first part of Theorem 4.2. By the use of formulas (3.6) and (4.9) one finds that

$$J_1^*(\alpha, \eta) = J^*(\alpha, \eta) + 2bH_{\sigma k}\alpha_k H_\sigma(\alpha, \eta).$$

It follows from this equation, the relation $H_{\sigma k} = 0$ $(k > r')$ and the definition of $H_\sigma(\alpha, \eta)$ that the value of $J_1^*(\alpha, \eta)$ determined by the variation α_h, $\eta_i = \bar{\eta}_{ik}\alpha_k$ is given by the formula

$$(4.15) \qquad J_1^*(\alpha, \bar{\eta}_k\alpha_k) = J^*(\alpha, \bar{\eta}_k\alpha_k) + 2bH_{\sigma\tau}H_{\sigma\nu}\alpha_\tau\alpha_\nu \qquad (\tau, \nu = 1, \cdots, r').$$

When the first r' of the α's are zero, the arc α_h, $\eta_i = \bar{\eta}_{ik}\alpha_k$ and the multipliers $\lambda_{\rho k}\alpha_k$, $\mu_{\gamma k}\alpha_k$ define an extremal of each of the accessory Mayer fields just constructed. One then has $J_1^*(\alpha, \bar{\eta}_k\alpha_k) = J(\alpha, \eta) > 0$ if $(\alpha) \neq (0)$, since the second variation J of I is positive along E. On the other hand the last term in (4.15) is a positive definite quadratic form in the first r' of the α's since the matrix $\| H_{\sigma\tau} \|$ $(\sigma, \tau = 1, \cdots, r')$ has rank r'. It follows from the theory of quadratic forms that if the constant b is chosen sufficiently large one has $J_1^*(\alpha, \bar{\eta}_k\alpha_k) > 0$ for all $(\alpha) \neq (0)$. Moreover, for an arbitrary arc $(\alpha, \eta) \neq (0, \eta)$ satisfying the conditions (2.9b) one has $\eta_i(x_s) = \bar{\eta}_{ik}(x_s)\alpha_k$ $(s = 1, 2)$ and hence $J_1^*(\alpha, \eta) = J_1^*(\alpha, \bar{\eta}_k\alpha_k) > 0$. The accessory extremals (4.14) accordingly have the properties described in Theorem 4.2 and the theorem is established.

5. **Proof of Theorem** 2.2. In the proof of the sufficiency conditions described in Theorem 2.2 we make use of the following lemma.

LEMMA 5.1. *In the proof of Theorem* 2.2 *one can assume without loss of generality that there exists for the extremal E a conjugate system* u_{ij}, v_{ij} *having* $| u_{ij}(x) | \neq 0$ *on* x_1x_2.

If one accepts for the moment the truth of this lemma one can prove Theorem 2.2 as follows: Select for E a set of r accessory extremals (4.1) having the properties described in Theorem 4.2 and let $\mathfrak{F}$ be a Mayer field related to E and these accessory extremals in the manner described in Theorem 4.1. By the use of Theorems 3.2 and 4.2 it is found that if one takes the neighborhood $\mathfrak{F}$ of E sufficiently small, the relation $I^*(C) > I^*(E)$ will hold for every admissible arc in $\mathfrak{F}$ satisfying the conditions (1.3b) and having its components a_h distinct from those belonging to E. Moreover, one can choose $\mathfrak{F}$ so small that the elements $[a, x, y, p(a, x, y), l(a), m(a, x, y)]$ of the field will lie in the neighborhood N

of the elements (a, x, y, y', l, m) on E prescribed by the Weierstrass condition II_N and on which the determinant (2.2) is different from zero. From the remarks following formula (2.6) one finds that the condition (3.3) holds in $\mathfrak{F}$. Theorem 2.2 now follows from Theorem 3.1.

Lemma 5.1 will be established by a transformation of our problem similar to that used by Denbow (IV). In the proof of Lemma 5.1 we can suppose that the functions $x_s(a)$ appearing in equations (1.3b) are constants since this result can be brought about by replacing x by a new variable t by means of the transformation $x = x_1(a) + t[x_2(a) - x_1(a)]$ $(0 \leqq t \leqq 1)$. Let δ be a constant related to E as described in Lemma 2.1. No generality is lost in assuming that $x_1 = 0$, $x_2 = q\delta$, where q is an integer. In fact we may suppose that $\delta = 1$ since this result can be brought about by setting $x = \delta t$. We then have $x_1 = 0$, $x_2 = q$. Let C be an admissible arc (1.2) satisfying the conditions (1.3) and denote by C_σ $(\sigma = 1, \cdots, q)$ the subarc of C determined by the interval $\sigma - 1 \leqq x \leqq \sigma$. The problem here studied, which we have designated as problem A, can be considered as the problem of minimizing the expression

$$I = g(a) + \int_{C_1} f\,dx + \cdots + \int_{C_q} f\,dx$$

in the class of admissible subarcs $C_1, \cdots, C_q$ such that

$$I_\rho = g_\rho(a) + \int_{C_1} f_\rho\,dx + \cdots + \int_{C_q} f_\rho\,dx = 0$$

and having the following properties: Each subarc satisfies the conditions $\varphi_\gamma = 0$; the final end point $(a, x, y) = (a, \tau, b_\tau)$ of C_τ $(\tau < q)$ is identical with the initial point of $C_{\tau+1}$; the initial point of C_1 is $(a, 0, y_1(a))$ and the final end point of C_q is $(a, q, y_2(a))$. If we map the subarc C_σ into an arc in atY_σ-space by the transformation

$$x = X_\sigma(t) = \sigma - 1 + t, \qquad y_i = Y_{i\sigma},$$

it is seen that the above formulation of problem A is equivalent to the problem of minimizing the expression

$$I = g(a) + \int_0^1 \{f(a, X_1(t), Y_1, Y_1') + \cdots + f(a, X_q(t), Y_q, Y_q')\}\,dt$$

in the class of admissible arcs

$$a_h, \qquad b_{i\tau}, \qquad Y_{i\sigma}(t) \qquad\qquad (0 \leqq t \leqq 1),$$

$$(h = 1, \cdots, r;\ i = 1, \cdots, n;\ \tau = 1, \cdots, q-1;\ \sigma = 1, \cdots, q)$$

satisfying the conditions

$$\varphi_\gamma(a, X_\sigma(t), Y_\sigma, Y_\sigma') = 0,$$

$$Y_{i1}(0) = y_{i1}(a), \qquad Y_{i\nu}(0) = b_{i\tau} = Y_{i\tau}(1), \qquad Y_{iq}(1) = y_{i2}(a) \qquad (\nu = \tau + 1),$$

$$I_\rho = g_\rho(a) + \int_0^1 \{f_\rho(a, X_1(t), Y_1, Y_1') + \cdots + f_\rho(a, X_q(t), Y_q, Y_q')\}\,dt = 0.$$

By considering subarcs, we can easily see that the extremal E with multipliers l_ρ, $m_\gamma(x)$ is transformed into an extremal E^* with multipliers l_ρ, $m_\gamma[X_\sigma(t)]$. The Weierstrass and non-singularity conditions for E^* are equivalent to those for the subarcs of E and hence to those for E itself. Moreover, the first and second variations of I along E are transformed into the first and second variations of the function I along E^*, admissible variations being transformed in the same manner as admissible arcs. Furthermore accessory extremals for E with discontinuities at integral values of x are transformed into accessory extremals for E^* without discontinuities. Finally by Lemma 2.1 and our choice of the subintervals $\sigma - 1 \leqq x \leqq \sigma$ there exist q conjugate systems u^σ_{ij}, v^σ_{ij} $(\sigma = 1, \cdots, q)$ such that for each σ the determinant $| u^\sigma_{ij}(x) |$ is different from zero on $\sigma - 1 \leqq x \leqq \sigma$. Let $U_{i\sigma, j\nu}(t)$, $V_{i\sigma, j\nu}(t)$ $(i, j = 1, \cdots, n; \sigma, \nu = 1, \cdots, q)$ be identically zero when $\sigma \neq \nu$ and equal to $u^\sigma_{ij}[X_\sigma(t)]$, $V^\sigma_{ij}[X_\sigma(t)]$ when $\sigma = \nu$. The functions so defined form a conjugate system for E^*, as one readily verifies. Moreover, the determinant $| U_{i\sigma, j\nu} |$ where the indices i, σ determine the rows and j, ν the columns, is equal to the product of the determinants $| u^\sigma_{ij}[X_\sigma(t)] |$ and is accordingly different from zero on $0 \leqq t \leqq 1$. This proves Lemma 5.1, and the proof of Theorem 2.2 is complete.

THEOREM 5.1. *Let E be an extremal satisfying the hypotheses of Theorem* 2.2 *and such that there exists for E a conjugate system u_{ij}, v_{ij} having $| u_{ij}(x) | \neq 0$ on x_1x_2. Then there exists a set of multipliers $l_\rho(a)$ such that E affords a proper strong relative minimum to the expression $I + l_\rho(a)I_\rho$ relative to neighboring admissible arcs satisfying the conditions* (1.3a) *and* (1.3b) *but not necessarily* (1.3c).

To prove this we choose $l_\rho(a)$ to be the multipliers of the Mayer field used in the proof of Theorem 2.2. By examining the proof of Theorems 2.2 and 3.1, we easily see that these multipliers have the property described in Theorem 5.1.

REFERENCES

I. G. A. BLISS, *The problem of Bolza in the calculus of variations*, Lectures delivered at the University of Chicago, Winter Quarter, 1935. See also Annals of Mathematics, vol. 33(1932), pp. 261–274.

II. G. A. BLISS, *Normality and abnormality in the calculus of variations*, Transactions of the American Mathematical Society, vol. 43(1938), pp. 365–376.

III. C. P. BRADY, *The minimum of a function of integrals in the calculus of variations*, Dissertation, The University of Chicago, 1938.

IV. C. H. DENBOW, *A generalized form of the problem of Bolza*, Contributions to the calculus of variations, 1933–1937, The University of Chicago, pp. 449–484.

V. W. L. DUREN, *Contractable problems of Bolza*, Bulletin of the American Mathematical Society, vol. 42(1936), p. 812, abstract no. 418, and vol. 44(1938), p. 32, abstract no. 17.

VI. M. R. HESTENES, *A direct sufficiency proof for the problem of Bolza in the calculus of variations*, Transactions of the American Mathematical Society, vol. 42(1937), pp. 141–153.

VII. M. R. HESTENES AND W. T. REID, *A note on the Weierstrass condition in the calculus of variations*, to appear in Bulletin of the American Mathematical Society, vol. 45 (1939).

VIII. MARSTON MORSE, *Sufficient conditions in the problem of Lagrange with variable end points*, American Journal of Mathematics, vol. 53(1931), pp. 517–546.

IX. W. T. Reid, *A direct expansion proof of sufficient conditions for the non-parametric problem of Bolza*, Transactions of the American Mathematical Society, vol. 42 (1937), pp. 183–190.

X. W. T. Reid, *Isoperimetric problems of Bolza in non-parametric form* (to be published soon).

Further references will be found in these papers.

The University of Chicago.

A GENERALIZED LAMBERT SERIES

By J. M. Dobbie

1. **Introduction.** In his important paper of 1913 Knopp [5][1] proposed to call any series of the form

$$\sum_{n=1}^{\infty} b_n x^n (1 - x^n)^{-1} \tag{1.1}$$

a Lambert series in honor of J. H. Lambert who in 1771 was the first to treat special series of this type [6].

The purpose of this paper is to discuss a more general series, namely:

$$\sum_{n=1}^{\infty} b_n x^{\lambda n} h(x^n), \tag{1.2}$$

where λ is a positive integer and $h(x)$ is a function of x which is analytic in the interior of the unit circle and which has a value different from zero at $x = 0$. In §§3 and 4 we suppose further that $h(x)$ has on the unit circle a finite number of singularities of which at least one is a pole.

In his paper Knopp proves the following theorem.[2]

Let the coefficients b_n of the series (1.1) be such that for a definite integer k all the k series

$$\sum_{\nu=1}^{\infty} \frac{b_{k\nu+l}}{k\nu + l} \qquad (l = 0, 1, 2, \cdots, k-1) \tag{1.3}$$

converge. Then if for such a k and for k' prime to k we write $x_0 = e^{2\pi i k'/k}$, we have for radial approach of x to x_0 the relation

$$\lim_{x \to x_0} \left\{ (1 - x/x_0) \sum_{n=1}^{\infty} b_n x^n (1 - x^n)^{-1} \right\} = \sum_{\nu=1}^{\infty} \frac{b_{k\nu}}{k\nu}.$$

From this theorem it follows that the function defined by the series (1.1) cannot be continued analytically across the unit circle if the hypotheses of the theorem are satisfied for an infinite number of values of k for each of which the series $\sum_{\nu=1}^{\infty} b_{k\nu}/(k\nu)$ satisfies the additional condition of having its sum different from zero.

Recently Mary Cleophas Garvin [2] obtained corresponding results for series of the form

$$\sum_{n=1}^{\infty} b_n x^{\lambda n} (1 - x^{\mu n})^{-1}, \tag{1.4}$$

which are obtained from our series in the case $h(x) = 1/(1 - x^{\mu})$.

Received November 16, 1938.

[1] The numbers in brackets refer to the list of references at the end of the paper.

[2] Theorem 3 of §2.

Later in the same year in which Knopp's paper appeared Hardy [3] showed that Knopp's theorem can be generalized by replacing the hypothesis of the convergence of the series (1.3) by the hypothesis that these series are summable Cesàro (C, p) for some non-negative integer p. He also showed that a part of this latter hypothesis could be replaced by a more general one, namely, that the p-th Cesàro sums $A^p_{\nu,l}$ formed from the series $\sum_{\nu=0}^{\infty} a_{\nu,l} = \sum_{\nu=0}^{\infty} b_{k\nu+l}$ $(l = 1, 2, \cdots, k-1)$ satisfy the relation $A^p_{\nu,l} = o(\nu^{p+1})$, the hypothesis on the series $\sum_{\nu=1}^{\infty} b_{k\nu}/(k\nu)$ and the conclusion remaining the same. He pointed out that this latter theorem, as well as Knopp's, is equivalent to two theorems, the second of which he states in more general form.

In §§3 and 4 of this paper we prove theorems for our generalized series (1.2) which correspond to these theorems of Hardy for ordinary Lambert series. The theorems of §§3 and 4 are the most important results of this paper.

2. Definition of the series. Convergence. Necessary and sufficient conditions for expanding a function in a series of this type. Let $h(t)$ be a function which is analytic in the interior of the unit circle and which has a value different from zero at $t = 0$. Write its power series expansion as

$$h(t) = \alpha_\lambda + \alpha_{\lambda+1}t + \alpha_{\lambda+2}t^2 + \cdots, \tag{2.1}$$

where $\alpha_\lambda \neq 0$ and λ is a positive integer. Then form the series

$$f(t) = \sum_{n=1}^{\infty} b_n t^{\lambda n} h(t^n). \tag{2.2}$$

This affords a generalization of the Lambert series (1.1) and the series (1.4) of Garvin.

Using well known methods, we can easily prove

THEOREM 2.1. *If the series $\sum_{n=1}^{\infty} b_n t^{\lambda n}$ converges in the interior of the circle $|t| = r$ (whence $r \leq 1$), then the series (2.2) converges absolutely and uniformly in every closed region in the interior of $|t| = r$. Moreover, expansions of the form (2.2) are unique.*

The uniqueness of the expansions follows from the hypothesis that $h(0) \neq 0$. For, if $f(t)$ has a second convergent expansion $\sum_{n=1}^{\infty} B_n t^{\lambda n} h(t^n)$ in the neighborhood of $t = 0$ and we equate the derivatives of order $k\lambda$ $(k = 1, 2, 3, \cdots)$ of the two expansions at $t = 0$, we get $B_k(k\lambda)!h(0) = b_k(k\lambda)!h(0)$. Since $h(0) \neq 0$, $B_k = b_k$ $(k = 1, 2, \cdots)$.

From this theorem we have that if the associated power series

$$\sum_{n=1}^{\infty} b_n t^{\lambda n} \tag{2.3}$$

has a radius of convergence different from zero, then the function $f(t)$ is analytic in the neighborhood of the origin and has a power series expansion of the form

$$f(t) = a_\lambda t^\lambda + a_{\lambda+1} t^{\lambda+1} + a_{\lambda+2} t^{\lambda+2} + \cdots. \tag{2.4}$$

From (2.1) and (2.2) we have

$$f(t) = \sum_{n=1}^{\infty} b_n (\alpha_\lambda t^{\lambda n} + \alpha_{\lambda+1} t^{(\lambda+1)n} + \alpha_{\lambda+2} t^{(\lambda+2)n} + \cdots). \tag{2.5}$$

Comparing (2.5) with (2.4) we see that

$$a_n = \sum_{d \mid n} b_d \alpha_{n/d}, \tag{2.6}$$

the sum running over all the divisors d of n, with the understanding that $\alpha_i = 0$, $i < \lambda$. In general, the system of equations (2.6) cannot be solved uniquely for the b's in terms of the a's. However, if $\lambda = 1$, this is possible, since in this case the equations (2.6) are consistent in the b's and the recurrence relation becomes

$$b_1 \alpha_1 = a_1; \qquad b_n \alpha_1 = a_n - \sum_{\substack{d \mid n \\ d \neq n}} b_d \alpha_{n/d} \qquad (n = 2, 3, 4, \cdots),$$

in which $\alpha_1 \neq 0$.

Hence, for every function analytic at zero and vanishing there, such as that in (2.4) (with $\lambda = 1$), there exists a formal transformation into a series of the form (2.2) (with $\lambda = 1$). It is easy to show that this series always converges in the neighborhood of $t = 0$. Thus, we have

THEOREM 2.2. *A necessary and sufficient condition that a function $f(t)$ shall admit a convergent expansion in the form $\sum_{n=1}^{\infty} b_n t^n h(t^n)$ is that $f(t)$ shall be analytic at $t = 0$ and shall vanish there.*

If, for general λ, we consider only those equations in (2.6) for which $n = k\lambda$ $(k = 1, 2, 3, \cdots)$, the resulting expansion $\sum_{n=1}^{\infty} b_n t^{\lambda n} h(t^n)$ converges in the neighborhood of $t = 0$ but does not necessarily represent $f(t)$ (as given in (2.4)) there, as it is certain that the derivatives of orders $0, 1, 2, \cdots, \lambda - 1, \lambda, 2\lambda, 3\lambda, \cdots$ only coincide at $t = 0$. However, if $h(t)$ and $f(t)$ are such that $a_n = \alpha_n = 0$ for all values of n which are not positive integral multiples of λ, we get

THEOREM 2.3. *A necessary and sufficient condition that a function $f(t^\lambda)$ shall admit a convergent expansion in the form $\sum_{n=1}^{\infty} b_n t^{\lambda n} h(t^{\lambda n})$ is that $f(t)$ shall be analytic at $t = 0$ and shall vanish there.*

From this last theorem it is seen that we may confine our attention to the case $\lambda = 1$ in seeking the solution of the system (2.6) for series of the form $\sum_{n=1}^{\infty} b_n t^{\lambda n} h(t^{\lambda n})$. As a special case of (2.6) consider

$$\sum_{d|n} \beta_d \alpha_{n/d} = \begin{cases} 1 & \text{for } n = 1, \\ 0 & \text{otherwise,} \end{cases} \qquad \alpha_1 \beta_1 = 1, \tag{2.7}$$

which we know has a solution which gives rise to

$$t = \sum_{\nu=1}^{\infty} \beta_\nu t^\nu h(t^\nu).$$

From this equation we have

$$\sum_{\mu=1}^{\infty} a_\mu t^\mu = \sum_{\mu=1}^{\infty} \sum_{\nu=1}^{\infty} a_\mu \beta_\nu t^{\mu\nu} h(t^{\mu\nu}) = \sum_{m=1}^{\infty} \sum_{d|m} a_d \beta_{m/d} t^m h(t^m), \tag{2.8}$$

the interchange of operations being justified by Theorem 2.1. From equation (2.8) and equation (2.2) (with $\lambda = 1$), we get

$$b_n = \sum_{d|n} a_d \beta_{n/d},$$

and this gives an effective solution of (2.6) provided we can solve (2.7) for the β's in terms of the α's.

3. **Some theorems on limits**. As a basis for the discussion of the existence of a natural boundary for a function defined by a generalized Lambert series of the form (2.2), we prove two theorems on limits in this section. These theorems are generalizations of Theorems 2 and 3 of Hardy [3] for ordinary Lambert series.

THEOREM 3.1. *Let $h(y)$ be a function which is analytic in the interior of the unit circle and which has on the unit circle a finite number of singularities of which at least one is a pole,*[3] *say at $y = 1$ and of order r. Let the series $\sum_{n=1}^{\infty} a_n$ be summable (C, p) with sum S and let λ be a positive integer. Then, if $y \to 1$ through real values less than 1, we have*

$$\lim_{y\to 1} \sum_{n=1}^{\infty} a_n n^r y^{\lambda n} (1 - y)^r h(y^n) = SA, \tag{3.1}$$

where A is the limit of $(1 - y)^r h(y)$ as $y \to 1$.

In the proof of this theorem we shall need a theorem proved by Bromwich [1] in 1908. For our purpose we state the theorem in the following form:

[3] If the pole is at $z = e^{2\pi i\alpha}$, apply the rotation $z = e^{2\pi i\alpha} y$.

If $\sum a_n$ is summable (C, p) with sum S and $f_n(y)$ is a function of y such that

$$\left.\begin{aligned} &(1) && \sum n^p \,|\, \Delta^{p+1} f_n(y) \,| < K \quad \text{(independent of } y\text{)} \\ &(2) && \lim_{n \to \infty} n^p f_n(y) = 0 \end{aligned}\right\} \quad 0 \leqq y < 1,$$

$$(3) \qquad \lim_{y \to 1} f_n(y) = A \quad \text{(independent of } n\text{)},$$

then $\sum a_n f_n(y)$ converges for $0 \leqq y < 1$, and $\lim\limits_{y \to 1} \sum a_n f_n(y) = SA$.

To apply this theorem to the limit in Theorem 3.1 above, let

$$f_n(y) = n^r y^{\lambda n} (1 - y)^r h(y^n) = n^r y^{\lambda n} (1 + y + y^2 + \cdots + y^{n-1})^{-r} g(y^n),$$

where $g(y) = (1 - y)^r h(y)$. Then condition (3) of Bromwich's theorem is satisfied for $\lim\limits_{y \to 1} f_n(y) = \lim\limits_{y \to 1} g(y) = A$, which is independent of n. Also, it is easy to show that condition (2) is satisfied.

To show that condition (1) is satisfied we first show that $\Delta^{p+1} f_n(y)$ can be written

$$(3.2) \qquad \Delta^{p+1} f_n(y) = (1 - y)^{p+1} y^{\lambda n} f_{n,p+1}(y),$$

where

$$(3.3) \qquad | f_{n,p+1}(y) | < B \qquad \text{(independent of } y \text{ for } 0 \leqq y \leqq 1\text{)}.$$

This involves showing that $\lim\limits_{y \to 1} f_{n,p}(y)$ is finite and independent of n, where

$$f_{n,p}(y) = (1 - y)^{-p} \sum_{s=0}^{p} (-1)^s \binom{p}{s} (n + s)^r (1 + y + y^2 + \cdots + y^{n+s-1})^{-r} g(y^{n+s}).$$

The proof of this statement is elementary but a bit tedious and is omitted here.

If we use equation (3.2) with condition (3.3), condition (1) is implied by

$$(3.4) \qquad \sum_{n=1}^{\infty} n^p \,|\, \Delta^{p+1} f_n(y) \,| < B(1 - y)^{p+1} \sum_{n=1}^{\infty} n^p y^{\lambda n},$$

where B is independent of y for $0 \leqq y \leqq 1$. But $\sum\limits_{n=1}^{\infty} n^p y^{\lambda n}$ has a pole of order $p + 1$ at $y = 1$ as its only singularity in the interval $0 \leqq y \leqq 1$ as is seen by inspection and simple induction from the relation

$$\sum_{n=1}^{\infty} n^p y^n = y \frac{d}{dy} \cdot y \frac{d}{dy} \cdot \cdots \cdot y \frac{d}{dy} \frac{1}{1 - y},$$

in which the operator $y \dfrac{d}{dy}$ is applied p times. Hence, from (3.4) it follows that there exists a constant K (independent of y for $0 \leqq y \leqq 1$) such that $\sum\limits_{n=1}^{\infty} n^p \,|\, \Delta^{p+1} f_n(y) \,| < K$.

Therefore, Bromwich's theorem can be applied to show that the limit in (3.1) is SA.

THEOREM 3.2. *Let $h(x)$ be defined as in Theorem 3.1 and let α be any point inside or on the unit circle which is not an essential singularity of $h(x)$. Write $h(x) = (\alpha - x)^\sigma g(x)$, where $g(x)$ is analytic at $x = \alpha$ and does not vanish there. Let l be any integer of the set $1, 2, 3, \cdots, k-1$. Form the series*

$$F(\rho) = \sum_{\nu=0}^{\infty} b_{k\nu+l}\,\rho^{\lambda(k\nu+l)}\,h(\alpha\rho^{k\nu+l})$$

and let the coefficients $b_{k\nu+l}$ be such that the p-th Cesàro sum formed from the series $\sum_{\nu=0}^{\infty}(k\nu + l)^\sigma b_{k\nu+l}$ satisfies the relation $B_{\nu,l}^p = o(\nu^{p+1})$. Then $\lim_{\rho\to 1}(1-\rho)^{1-\sigma}F(\rho) = 0$.

We can write $\lim (1-\rho)^{1-\sigma}F(\rho)$ as

$$\alpha^\sigma \lim_{\rho\to 1}(1-\rho)\sum_{\nu=0}^{\infty}(k\nu+l)^\sigma b_{k\nu+l}(\nu + l/k)^{-\sigma}\rho^{\lambda k\nu}(1-\rho^{k\nu+l})^\sigma(1-\rho^k)^{-\sigma}g(\alpha\rho^{k\nu+l}).$$

In the series above let $B_{\nu,l} = (k\nu + l)^\sigma b_{k\nu+l}$ and let

$$f_{\nu,l}(\rho) = (\nu + l/k)^{-\sigma}\rho^{\lambda k\nu}(1-\rho^{k\nu+l})^\sigma(1-\rho^k)^{-\sigma}g(\alpha\rho^{k\nu+l}).$$

An argument similar to that used in Theorem 3.1 can be made to show that $\Delta^{p+1}f_{\nu,l}(\rho) = (1-\rho)^{p+1}\rho^{\lambda k\nu}O(1)$. Since $\sum B_{\nu,l}f_{\nu,l}(\rho) = \sum B_{\nu,l}^p\Delta^{p+1}f_{\nu,l}(\rho)$ and by hypothesis $B_{\nu,l}^p = o(\nu^{p+1})$, the theorem follows from the latter part of Hardy's proof of his Theorem 3.[4]

4. Existence of a natural boundary.

THEOREM 4.1. *Let $h(x)$ be defined as in Theorem 3.1. Take k and k' as any two relatively prime positive integers and write $x_0 = e^{2\pi i k'/k}$. Let the points x_0^l $(l = 1, 2, \cdots, k-1)$ be poles of order $r_l \leqq r - 1$ of $h(x)$. For λ a positive integer form the series*

$$f(x) = \sum_{n=1}^{\infty} b_n x^{\lambda n} h(x^n), \tag{4.1}$$

and let the coefficients b_n be such that

(i) *the series $\sum_{\nu=1}^{\infty}(k\nu)^{-r}b_{k\nu}$ is summable (C, p) with sum S, and either*

(ii) *the series $\sum_{\nu=0}^{\infty}(k\nu + l)^{-r_l-1}b_{k\nu+l}$ $(l = 1, 2, \cdots, k-1)$ are summable (C, p),*

or

[4] Hardy [3], pp. 196–197.

(ii′)[5] *the p-th Cesàro sums* $B^p_{\nu,l}$ *formed from the series* $\sum_{\nu=0}^{\infty} (k\nu + l)^{-r_l} b_{k\nu+l}$ $(l = 1, 2, \cdots, k - 1)$ *are such that* $B^p_{\nu,l} = o(\nu^{p+1})$. *Then, for radial approach of* x *to* x_0, *we have the relation*

$$\lim_{x\to x_0} (1 - x/x_0)^r f(x) = SA,$$

where A *is the limit of* $(1 - x)^r h(x)$ *as* $x \to 1$. *If this is true for an infinite number of integers* k, *for each of which* $S = \sum_{\nu=1}^{\infty} (k\nu)^{-r} b_{k\nu} \neq 0$, *then* $f(x)$ *cannot be continued analytically across the unit circle.*

Write $x = \rho e^{2\pi i k'/k}$. Then we must determine

$$\lim_{\rho\to 1} (1 - \rho)^r \sum_{n=1}^{\infty} b_n x^{\lambda n} h(x^n). \tag{4.2}$$

First, consider those terms in the series in (4.2) for which n is a multiple of k: $n = k\nu$. Then $x^k = \rho^k$ and the contribution to the limit in (4.2) for such values of n becomes

$$\begin{aligned}\lim_{\rho\to 1} (1 - \rho)^r \sum_{\nu=1}^{\infty} b_{k\nu} x^{\lambda k\nu} h(x^{k\nu}) &= \lim_{\rho\to 1} \left(\frac{1-\rho}{1-\rho^k}\right)^r \sum_{\nu=1}^{\infty} b_{k\nu} \rho^{\lambda k\nu} (1 - \rho^k)^r h(\rho^{k\nu}) \\ &= k^{-r} \lim_{\rho\to 1} \sum_{\nu=1}^{\infty} b_{k\nu} \rho^{\lambda k\nu} (1 - \rho^k)^r h(\rho^{k\nu}).\end{aligned}$$

Let $y = \rho^k$ and this limit becomes $\lim_{\rho\to 1} \sum_{\nu=1}^{\infty} \frac{b_{k\nu}}{(k\nu)^r} \nu^r y^{\lambda\nu} (1 - y)^r h(y^\nu) = SA$ by Theorem 3.1.

To complete the proof of Theorem 4.1 we must show that the contribution to the limit in (4.2) of those terms for which $n = k\nu + l$ $(l = 1, 2, 3, \cdots, k - 1)$ is zero. We must consider

$$\begin{aligned}&\lim_{x\to x_0} (1 - x/x_0)^r \sum_{\nu=0}^{\infty} b_{k\nu+l} x^{\lambda(k\nu+l)} h(x^{k\nu+l}) \\ &\qquad = x_0^{\lambda l} \lim_{\rho\to 1} (1 - \rho)^{r-r_l-1} (1 - \rho)^{1+r_l} \sum_{\nu=0}^{\infty} b_{k\nu+l} \rho^{\lambda(k\nu+l)} h(x_0^l \rho^{k\nu+l}).\end{aligned}$$

This limit is zero as we see from Theorem 3.2 by putting $x_0^l = \alpha$, $r_l = -\sigma$, noting the hypothesis $r - r_l \geqq 1$.

The statement concerning analytic continuation follows from the condition that $SA \neq 0$ together with the fact that the points $x_0 = e^{2\pi i k'/k}$ form an everywhere dense set on the circumference of the unit circle.

[5] That condition (ii′) is slightly more general than condition (ii) is proved by Hardy and Littlewood [4], p. 435, Theorem 14.

References

1. T. J. I'A. Bromwich, Mathematische Annalen, vol. 65(1908), pp. 350–369.
2. M. C. Garvin, American Journal of Mathematics, vol. 58(1936), pp. 507–513.
3. G. H. Hardy, Proceedings of the London Mathematical Society, (2), vol. 13(1913), pp. 192–198.
4. G. H. Hardy and J. E. Littlewood, Proceedings of the London Mathematical Society, (2), vol. 11(1912), pp. 411–478.
5. Konrad Knopp, Journal für die reine und angewandte Mathematik, vol. 142(1913), pp. 283–315.
6. J. H. Lambert, *Anlage zur Architectonik, oder Theorie des Einfachen und Ersten in der philosophischen und mathematischen Erkenntnis*, vol. II, 1771, p. 507 (§875).

University of Illinois.

PERSYMMETRIC AND JACOBI DETERMINANT EXPRESSIONS FOR ORTHOGONAL POLYNOMIALS

By Vivian Eberle Spencer

Introduction. Work with orthogonal Tchebycheff polynomials (OP) has usually taken as its point of attack the notion of a weight function and the corresponding moment problem. In the study of OP two important expressions for them as determinants arise. The first, or *persymmetric determinant expression*, is obtained by replacing the elements of the last row of a certain positive persymmetric determinant by powers of x; the second, or *Jacobi determinant expression*, results when the characteristic determinant of a certain Jacobi matrix is written. The present paper undertakes a study of orthogonal and related persymmetric polynomials from the standpoint of the theory of matrices and determinants. In all fundamental theory the notion of a weight function is entirely avoided.

The persymmetric determinant expression leads to a classification of sequences of these polynomials into sets S, each of which is found to contain one and only one symmetric sequence. Properties of sets S are investigated. In considering the Jacobi determinant expression we come upon certain finite sequences[1] $\{\Theta_i(x)\}_1^n$ of orthogonal polynomials which associate themselves in a very simple manner with any sequence of OP $\{\Phi_n(x)\}$. The study of $\{\Theta_i(x)\}_1^n$ leads to new bounds for the zeros of $\Phi_n(x)$. Properties of the continued fraction associated with $\{\Theta_i(x)\}_1^n$ are obtained. Combining these results, we are led to a set of theorems regarding a formally defined interval of orthogonality for $\{\Phi_n(x)\}$. A theorem of Krein[2] has important bearing on our study. We close with its extension. Throughout the paper applications are made to the classical orthogonal polynomials.

For OP we use the notation adopted by J. Shohat.[3]

1. A classification of sequences of persymmetric polynomials.

A *persymmetric or Hankel determinant* is a determinant in which each line perpendicular to the

Received December 20, 1938; section 1 of this paper was presented to the American Mathematical Society, September 13, 1935. The author wishes to express her gratitude to Professor J. Shohat for his highly-valued interest and suggestions throughout the preparation of this paper.

[1] By the notation $\{q_i\}_1^n$ will be understood the sequence $\{q_i\}$ $(i = 1, 2, \cdots, n)$.

[2] M. Krein, *Über das Spektrum der Jacobischen Form in Verbindung mit der Theorie der Torsionsschwingungen von Walzen* (in Russian), Rec. Math. Moscou, vol. 40(1933), pp. 455–465.

[3] J. Shohat, *Théorie Générale des Polynomes Orthogonaux de Tchebichef*, Mémorial des Sciences Math., Fasc. 66(1934).

principal diagonal has all its elements alike. Thus,

$$(1)\qquad \Delta_n = \begin{vmatrix} \alpha_0 & \alpha_1 & \cdots & \alpha_{n-1} \\ \alpha_1 & \alpha_2 & \cdots & \alpha_n \\ \cdots & \cdots & \cdots & \cdots \\ \alpha_{n-1} & \alpha_n & \cdots & \alpha_{2n-2} \end{vmatrix} \equiv [\alpha_{i+j}]_0^{n-1} \qquad (\Delta_0 = 1, \Delta_1 = \alpha_0)$$

is a persymmetric determinant of order n. Δ_n is determined when the $2n - 1$ elements of the principal and one adjacent minor diagonal are given. *We shall assume only that* $\Delta_n \neq 0$ $(n = 1, 2, \cdots)$, a broader assumption than $\Delta_n > 0$, which leads to OP.[4]

Consider the sequence of polynomials $\{\Phi_n(x)\}$, where

$$(2)\qquad \Phi_n(x) = \frac{1}{\Delta_n}\begin{vmatrix} \alpha_0 & \alpha_1 & \cdots & \alpha_n \\ \alpha_1 & \alpha_2 & \cdots & \alpha_{n+1} \\ \cdots & \cdots & \cdots & \cdots \\ \alpha_{n-1} & \alpha_n & \cdots & \alpha_{2n-1} \\ 1 & x & \cdots & x^n \end{vmatrix} \qquad (n = 1, 2, \cdots; \Phi_0 = 1).$$

By elementary transformations (2) may be reduced to the form

$$(3)\qquad \Phi_n(x) = \frac{(-1)^n}{\Delta_n}\begin{vmatrix} \alpha_1 - x\alpha_0 & \alpha_2 - x\alpha_1 & \cdots & \alpha_n - x\alpha_{n-1} \\ \alpha_2 - x\alpha_1 & \alpha_3 - x\alpha_2 & \cdots & \alpha_{n+1} - x\alpha_n \\ \cdots & \cdots & \cdots & \cdots \\ \alpha_n - x\alpha_{n-1} & \alpha_{n+1} - x\alpha_n & \cdots & \alpha_{2n-1} - x\alpha_{2n-2} \end{vmatrix},$$

a multiple of a persymmetric determinant. Hence, we shall call the $\Phi_n(x)$ *persymmetric polynomials.*

In substance Jacobi[5] first proved that the polynomials of $\{\Phi_n(x)\}$ satisfy the recurrence relation

$$(4)\qquad \Phi_n(x) = \left(x - \frac{\Delta_n''}{\Delta_n} + \frac{\Delta_{n-1}''}{\Delta_{n-1}}\right)\Phi_{n-1}(x) - \frac{\Delta_{n-2}\Delta_n}{\Delta_{n-1}^2}\Phi_{n-2}(x) \qquad (n = 2, 3, \cdots),$$

where

$$(5)\qquad \Delta_n'' = \begin{vmatrix} \alpha_0 & \alpha_1 & \cdots & \alpha_{n-2} & \alpha_n \\ \alpha_1 & \alpha_2 & \cdots & \alpha_{n-1} & \alpha_{n+1} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \alpha_{n-2} & \alpha_{n-1} & \cdots & \alpha_{2n-4} & \alpha_{2n-2} \\ \alpha_{n-1} & \alpha_n & \cdots & \alpha_{2n-3} & \alpha_{2n-1} \end{vmatrix} \qquad (n = 2, 3, \cdots; \Delta_0'' = 0, \Delta_1'' = \alpha_1).$$

[4] Shohat has shown that polynomials (2) with the broader condition $\Delta_n \neq 0$ lead to a generalization of orthogonal polynomials (Comptes Rendus, vol. 207(1938), pp. 556–558).

[5] C. G. J. Jacobi, *De eliminatione variabilis e duabus aequationibus algebraicis*, Journal für Mathematik, vol. 15(1836), pp. 101–124.

A direct derivation of (4) from (2) may be carried through, for example, by the method of Muir and Metzler.[6]

Write

$$
(6)\qquad \begin{cases} c_n = \dfrac{\Delta_n''}{\Delta_n} - \dfrac{\Delta_{n-1}''}{\Delta_{n-1}} & (n = 1, 2, \cdots), \\[2ex] \lambda_n = \dfrac{\Delta_{n-2}\Delta_n}{\Delta_{n-1}^2} & (n = 2, 3, \cdots; \lambda_1 = \alpha_0 = \Delta_1). \end{cases}
$$

Then (4) may be written as

$$
(7)\qquad \begin{aligned} &\Phi_n(x) = (x - c_n)\Phi_{n-1}(x) - \lambda_n \Phi_{n-2}(x) \\ &\qquad\qquad (n = 2, 3, \cdots; \Phi_0(x) = 1, \Phi_1(x) = x - c_1). \end{aligned}
$$

For all $\lambda_n > 0$, (7) is known to be characteristic for OP.

Given any sequence of numbers $\{\alpha_n\}$, an associated sequence of determinants $\{\Delta_n\}$ and an associated sequence of polynomials $\{\Phi_n(x)\}$ are determined. Denote the sequence of numerical values of the determinants Δ_n by $\{\delta_n\}$. Let us investigate the properties of sets of sequences $\{\Phi_n(x)\}$ with each of which is associated the same sequence of numbers $\{\delta_n\}$ by allowing the α_n to vary in such a way that $\{\delta_n\}$ remains fixed.

Definition. A *set* S is a set of sequences $\{\Phi_n(x)\}$ with each of which is associated the same sequence of numbers $\{\delta_n\}$.

The α_n will be called *moments* associated with $\{\Phi_n(x)\}$, and $\{\alpha_n\}$ will be said to generate $\{\Phi_n(x)\}$.

Consider any two such sequences $\{{}_i\Phi_n(x)\}$ and $\{{}_k\Phi_n(x)\}$ of the same set S to which correspond the sequences of determinants $\{{}_i\Delta_n\}$ and $\{{}_k\Delta_n\}$. $\{{}_k\Delta_n\}$ is obtainable from $\{{}_i\Delta_n\}$ by multiplying the matrices of $\{{}_i\Delta_n\}$ by matrices whose elements may or may not depend on $\{{}_i\alpha_n\}$. Obviously, the numerical value of the product of these matrix multipliers must be one. We may write the product of all matrix multipliers on the left and right of ${}_i\Delta_n$ respectively as $q^{-1}{}_1T_n$ and qT_n, where q is a constant ($\neq 0$) and ${}_1T_n$ and T_n are unit matrices. Moreover, we may assume ${}_1T_n$ and T_n reduced to the form

$$
(8)\qquad \left\| \begin{matrix} 1 & a_{12} & a_{13} & \cdots & a_{1n} \\ 0 & 1 & a_{23} & \cdots & a_{2n} \\ 0 & 0 & 1 & \cdots & a_{3n} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & \cdots & 1 \end{matrix} \right\| .
$$

For the first row and first column of ${}_1T_n\, {}_i\Delta_n\, T_n$ to be identical, it is necessary that ${}_1T_n \equiv T_n'$, the transpose of T_n. Hence, the elements of the sequence of

[6] Muir and Metzler, *Theory of Determinants*, 1930, p. 433.

determinants associated with any two sequences $\{{}_i\Phi_n(x)\}$ and $\{{}_k\Phi_n(x)\}$ of the same set S satisfy the relation

$$ {}_k\Delta_n = T'_n \, {}_i\Delta_n \, T_n \qquad (T_n \text{ given by (8)}). \tag{9} $$

T_n shall be said to carry $\{{}_i\Phi_n(x)\}$ into $\{{}_k\Phi_n(x)\}$.

Consider first the case where the a_{ij} in (8) do not depend on $\{\alpha_n\}$.

THEOREM I. *Let $\{{}_i\Phi_n(x)\}$ and $\{{}_k\Phi_n(x)\}$ be two sequences of a given set S. If the sequence of matrices $\{T_n\}$, independent of $\{\alpha_n\}$, carries $\{{}_i\Phi_n(x)\}$ into $\{{}_k\Phi_n(x)\}$, then*

$$ T_n = \left\| \begin{matrix} 1 & m & m^2 & m^3 & \cdots \\ 0 & 1 & 2m & 3m^2 & \cdots \\ 0 & 0 & 1 & 3m & \cdots \\ 0 & 0 & 0 & 1 & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots \end{matrix} \right\| \qquad (m \text{ arbitrary};\ n = 1, 2, \cdots). \tag{10} $$

Moreover, if $\{{}_i\Phi_n(x)\}$ and $\{{}_k\Phi_n(x)\}$ are so related, then, if the sequence of moments associated with $\{{}_i\Phi_n(x)\}$ is $\{\alpha_n\}$, the sequence of moments associated with $\{{}_k\Phi_n(x)\}$ is

$$ \left\{ m^n\alpha_0 + nm^{n-1}\alpha_1 + \frac{n(n-1)}{2!} m^{n-2}\alpha_2 + \cdots + \alpha_n \right\} \qquad (n = 0, 1, \cdots). $$

Proof. Sufficiency. If the right member product in (9) be formed with (10), the resulting left member is the determinant[7]

$$ \left[m^{i+j}\alpha_0 + (i+j)m^{i+j-1}\alpha_1 + \frac{(i+j)(i+j-1)}{2!} m^{i+j-2}\alpha_2 + \cdots + \alpha_{i+j} \right]_0^{n-1}. $$

Necessity. Assuming the a_{ij} of (8) independent of $\{\alpha_n\}$, form $T'_n \Delta_n T_n \equiv \| A_{ij} \|$. This is found to be

$$ \left\| \begin{matrix} \alpha_0 & a_{12}\alpha_0 + \alpha_1 & \cdots \\ a_{12}\alpha_0 + \alpha_1 & a_{12}^2\alpha_0 + 2a_{12}\alpha_1 + \alpha_2 & \cdots \\ a_{13}\alpha_0 + a_{23}\alpha_1 + \alpha_2 & a_{12}a_{13}\alpha_0 + (a_{12}a_{23} + a_{13})\alpha_1 + (a_{12} + a_{23})\alpha_2 + \alpha_3 & \cdots \\ \cdots & \cdots & \cdots \end{matrix} \right\|, \tag{11} $$

where, since $\Delta_n = \| \alpha_{i+j-2} \|$, $T_n = \| a_{ij} \|$, $T'_n = \| a_{ji} \|$,

$$ A_{ij} = \sum_k a_{ki} \sum_l \alpha_{k+l-2} a_{lj} = \sum_k \sum_l a_{ki} a_{lj} \alpha_{k+l-2}; \tag{12} $$

but if (11) is persymmetric, $A_{ij} = A_{i+1,j-1}$, or

$$ A_{ij} = \sum_k \sum_l a_{ki} a_{lj} \alpha_{k+l-2} = \sum_k \sum_l a_{k,i+1} a_{l,j-1} \alpha_{k+l-2}. \tag{13} $$

Now compare coefficients, remembering that $a_{ii} = 1$ and $a_{ij} = 0$ if $i > j$.

[7] T. Muir, *On a property of persymmetric determinants*, Messenger of Math., vol. 11(1881), pp. 65–67.

THEOREM II. *$\{T_n\}$ defined by* (10) *carries any $\{\Phi_n(x)\}$ into $\{\Phi_n(x - m)\}$.*

Proof. By elementary transformations it is readily shown that $\Delta_n \Phi_n(x - m)$ is given by

$$(14)\qquad \begin{vmatrix} \alpha_0 & m\alpha_0 + \alpha_1 & m^2\alpha_0 + 2m\alpha_1 + \alpha_2 & \cdots & \cdots \\ m\alpha_0 + \alpha_1 & m^2\alpha_0 + 2m\alpha_1 + \alpha_2 & \cdots & \cdots & \cdots \\ m^2\alpha_0 + 2m\alpha_1 + \alpha_2 & \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 1 & x & x^2 & \cdots & x^n \end{vmatrix}.$$

COROLLARY. *If a set S contains a certain sequence $\{\Phi_n(x)\}$, it also contains any sequence obtainable from $\{\Phi_n(x)\}$ by an arbitrary displacement along the x-axis. Moreover, all sequences associated by means of a matrix of type* (10) *are so related.*[8]

Let us now admit matrices (8) dependent upon $\{\alpha_n\}$.

THEOREM III. *Any given set S always contains a sequence $\{{}_l\Phi_n(x)\}$ generated by a sequence of moments $\{{}_l\alpha_n\}$ whose odd elements have arbitrarily preassigned values. Moreover, if the odd elements are given, the even ones are uniquely determined.*

Proof. Write the product

$$(15)\qquad T_n' \Delta_n T_n \equiv \| A_{ij} \| \equiv {}_l\Delta_n ,$$

where A_{ij} is the linear function of $\{\alpha_k\}_0^{i+j-2}$ indicated in (13). We shall need

LEMMA 1. *The determinant of coefficients of the system*

$$(16)\qquad A_{1n} = A_{2,n-1},\quad A_{2n} = A_{3,n-1},\quad \cdots,\quad A_{n-2,n} = A_{n-1,n-1},\quad A_{n-1,n} = K_{n-1}$$

(*where K_{n-1} is a constant*), *considered as a function of $a_{1n}, a_{2n}, a_{3n}, \cdots, a_{n-1,n}$, is Δ_{n-1}.*

Proof. Let $L = L(t_1, t_2, \cdots, t_n)$ denote a linear function of $t_1, t_2, \cdots, t_n$ of the form $\sum_1^{n-1} b_i t_i + t_n$. Since the $A_{i,n-1}$ are independent of $a_{1n}, a_{2n}, \cdots,$

[8] If all $\delta_n > 0$, then the moment problem $\{\alpha_n\}_0^\infty$ has at least one solution, i.e., there exists a function $\psi(x)$, bounded and non-decreasing in $(-\infty, \infty)$, such that $\int_{-\infty}^{\infty} x^n \, d\psi(x) = \alpha_n$ $(n = 0, 1, \cdots)$. Here the new moments $\{m^n\alpha_0 + nm^{n-1}\alpha_1 + \cdots + \alpha_n\}$ are evidently given by $\int_{-\infty}^{\infty} (x + m)^n d\psi(x) = \int_{-\infty}^{\infty} x^n d\psi(x - m)$. Hence, in this case transforming Δ_n by means of (10) corresponds to a displacement of $\{\Phi_n(x)\}$ along the x-axis, and the invariance of $\{\delta_n\}$ is evident.

$a_{n-1,n}$, write $A_{i,n-1} = C_i$. Then, by (11), letting C_i' denote the sum of the terms of A_{in} independent of $a_{1n}, a_{2n}, \cdots, a_{n-1,n}$, we may write (15) as

$$
\begin{aligned}
&L_1(\alpha_0)a_{1n} + L_1(\alpha_1)a_{2n} + \cdots + L_1(\alpha_{n-2})a_{n-1,n} = C_2 - C_1',\\
&L_2(\alpha_0, \alpha_1)a_{1n} + L_2(\alpha_1, \alpha_2)a_{2n} + \cdots + L_2(\alpha_{n-2}, \alpha_{n-1})a_{n-1,n} = C_3 - C_2',\\
&\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\\
&L_{n-1}(\alpha_0, \alpha_1, \cdots, \alpha_{n-2})a_{1n} + L_{n-1}(\alpha_1, \alpha_2, \cdots, \alpha_{n-1})a_{2n} + \cdots\\
&\qquad + L_{n-1}(\alpha_{n-2}, \alpha_{n-1}, \cdots, \alpha_{2n-4})a_{n-1,n} = K_{n-1} - C_{n-1}',
\end{aligned}
\tag{17}
$$

where L_i denotes a particular linear function of type L. The determinant of the coefficients of the linear system (17) reduces by elementary transformations to Δ_{n-1}.

Returning now to the proof of Theorem III, assume the a_{ij} of (8) are dependent on $\{\alpha_i\}$. Let the arbitrary preassigned sequence of values for the odd elements of $\{{}_\iota\alpha_n\}$ be denoted by $\{K_i\}$ $(i = 1, 2, \cdots)$. First, consider (15) for $n = 2$. From (11) we see that $T_2'\Delta_2 T_2$ is a function of a_{12} alone. Our conditions give $A_{12} = A_{21} = K_1$, a linear equation in a_{12} with the coefficient of a_{12} different from zero. Moreover, if we solve the latter, A_{22} is uniquely determined. Hence, the theorem is true for ${}_\iota\alpha_1$ and ${}_\iota\alpha_2$. Assume it true for $\{{}_\iota\alpha_i\}_0^{2n-4}$, that is, for every ${}_\iota\alpha_i$ in ${}_\iota\Delta_{n-1}$. Then we may consider ${}_\iota\Delta_n$ as a function of the $n-1$ unknowns $a_{1n}, a_{2n}, a_{3n}, \cdots, a_{n-1,n}$, these to be determined by the conditions that ${}_\iota\Delta_n$ be persymmetric and that ${}_\iota\alpha_{2n-3} = K_{n-1}$. From (11), ${}_\iota\Delta_n$ is symmetric; hence, our conditions lead to the $n-1$ linear equations (16) for these unknowns. Since we have made the fundamental assumption that no δ_i be zero, as an immediate consequence of Lemma 1, a unique solution of (16) for a_{1n}, $a_{2n}, a_{3n}, \cdots, a_{n-1,n}$ exists. Moreover, these a_{ij} determine A_{nn} uniquely. The theorem now follows by induction.

DEFINITION. A sequence of persymmetric polynomials $\{\Phi_n(x)\}$ is said to be a *symmetric sequence* if

$$\Phi_n(-x) \equiv (-1)^n\Phi_n(x) \qquad (n = 1, 2, \cdots), \tag{18}$$

that is, if

$$\Phi_n(x) = x^n + h_{n-2}x^{n-2} + h_{n-4}x^{n-4} + \cdots + h_1x, \quad \text{or } h_0. \tag{19}$$

LEMMA 2. *A necessary and sufficient condition that $\{\Phi_n(x)\}$ be a symmetric sequence is that all the odd moments $\{\alpha_i\}$ $(i = 1, 3, 5, \cdots)$ be zero.*

Proof. Sufficiency. By inspection of (2), under the assumption that $\alpha_i = 0$ $(i = 1, 3, \cdots)$, we see that the coefficient of x^{n-k} $(k = 1, 3, \cdots)$ is a determinant such that the column in which any non-zero element lies contains either one more or one less non-zero element than the row in which it lies. By suitable interchanges of rows and columns and expansion by Laplace's method, this determinant is seen to vanish.

Necessity. We have

$$\Phi_1(x) = \frac{1}{\alpha_0}\begin{vmatrix} \alpha_0 & \alpha_1 \\ 1 & x \end{vmatrix}. \tag{20}$$

Hence, if $\Phi_1(x)$ is symmetric, then $\alpha_1 = 0$. The proof may be completed by induction.

Combining this lemma with Theorem III, we have

COROLLARY 1. *A set S contains one and only one symmetric sequence.*

This sequence will be denoted by $\{{}_s\Phi_n(x)\}$ and the associated moments by $\{{}_s\alpha_n\}$.

COROLLARY 2. *Any given set S contains one and only one sequence $\{\Phi_n(x)\}$ for which $\{c_n\}$, the sequence of constants defined in* (6), *takes an arbitrarily preassigned set of values.*

Proof. From (6),

$$c_i = \frac{\Delta_i''}{\Delta_i} - \frac{\Delta_{i-1}''}{\Delta_{i-1}} = \frac{\alpha_{2i-1}\Delta_{i-1}}{\Delta_i} + \frac{1}{\Delta_i}\begin{vmatrix} \alpha_0 & \alpha_1 & \cdots & \alpha_{i-2} & \alpha_i \\ \alpha_1 & \alpha_2 & \cdots & \alpha_{i-1} & \alpha_{i+1} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \alpha_{i-1} & \alpha_i & \cdots & \alpha_{2i-3} & 0 \end{vmatrix} - \frac{\Delta_{i-1}''}{\Delta_{i-1}}. \tag{21}$$

Here Δ_{i-1}, $\Delta_i \neq 0$, and $\alpha_{2n-1}\Delta_{i-1}/\Delta_i$ is the only term containing α_{2i-1}; hence, c_i may be made to take any arbitrarily preassigned value by a suitable choice of α_{2i-1}, and, by Theorem III, the latter may be chosen arbitrarily. Moreover, by the same theorem the odd elements of $\{\alpha_n\}$ determine its even elements, hence the $\{c_i\}$, uniquely in S.

Remark. Corollary 1 may be obtained as an immediate consequence of Corollary 2, since the definition of a symmetric sequence, combined with the recurrence relation (7), leads at once to the conclusion that for such a sequence $c_n = 0$ $(n = 1, 2, \cdots)$.

Explicit expression for $\{{}_s\alpha_n\}$. For brevity write ${}_s\alpha_i \equiv \alpha_i$ $(i = 0, 1, \cdots)$, then $\alpha_{2i-1} = 0$. By a suitable interchange of rows and columns and Laplace expansion

$$\Delta_n = \begin{vmatrix} \alpha_0 & \alpha_2 & \cdots & \alpha_l \\ \alpha_2 & \alpha_4 & \cdots & \alpha_{l+2} \\ \cdots & \cdots & \cdots & \cdots \\ \alpha_l & \alpha_{l+2} & \cdots & \alpha_{2l} \end{vmatrix} \cdot \begin{vmatrix} \alpha_2 & \alpha_4 & \cdots & \alpha_m \\ \alpha_4 & \alpha_6 & \cdots & \alpha_{m+2} \\ \cdots & \cdots & \cdots & \cdots \\ \alpha_m & \alpha_{m+2} & \cdots & \alpha_{2m} \end{vmatrix}, \tag{22}$$

where l, m are respectively double the largest integer in $\frac{1}{2}(n-1)$ and $\frac{1}{2}n$. Successive application of (22) enables us to write α_n in terms of the $\{\delta_n\}$ determining

S, and the $\{\alpha_i\}_1^{n-1}$. Thus,

$$\alpha_2 = \frac{\delta_2}{\alpha_0}, \quad \alpha_4 = \frac{\delta_3}{\delta_2} + \frac{\alpha_2^2}{\alpha_0}, \quad \cdots,$$

$$\alpha_{2n-2} = \frac{\delta_n}{\delta_{n-1}} - \begin{vmatrix} \alpha_r & \alpha_{r+2} & \cdots & \alpha_m \\ \alpha_{r+2} & \alpha_{r+4} & \cdots & \alpha_{m+2} \\ \cdots & \cdots & \cdots & \cdots \\ \alpha_m & \alpha_{m+2} & \cdots & 0 \end{vmatrix} \div \begin{vmatrix} \alpha_r & \alpha_{r+2} & \cdots & \alpha_{m-2} \\ \alpha_{r+2} & \alpha_{r+4} & \cdots & \alpha_m \\ \cdots & \cdots & \cdots & \cdots \\ \alpha_{m-2} & \alpha_m & \cdots & \alpha_{2n-4} \end{vmatrix}, \tag{23}$$

where $r = 0$ for n odd and 2 for n even; and m is double the largest integer in $\frac{1}{2}n$.

Let us next investigate whether the even elements of $\{\alpha_n\}$ can be chosen arbitrarily.

THEOREM IV. *Associated with each set S there exists a set of intervals $\{\mathfrak{I}_i\}$ such that the set S contains a sequence $\{{}_I\Phi_n(x)\}$ generated by a sequence of moments $\{{}_I\alpha_n\}$ whose even elements $\{{}_I\alpha_i\}$ $(i = 2, 4, \cdots)$ have respectively values arbitrarily preassigned within the intervals $\{\mathfrak{I}_i\}$. Moreover, if the even moments are given, the odd moments are determined as two-valued functions of these even moments.*

Proof. The proof is analogous to that of Theorem III. Consider the arbitrary sequence of even moments $\{K_i\}$ $(i = 1, 2, \cdots)$, and the product $T'_n\Delta_n T_n$ written in the form (15). For $n = 2$ the conditions of our theorem require $A_{22} = K_1$, a quadratic equation in a_{12}. K_1 is restricted to values which will make the solution of this equation real. Hence,

$$K_1 \geqq \frac{\Delta_2}{\alpha_0} \equiv \frac{\delta_2}{\alpha_0},$$

and $\mathfrak{I}_1 \equiv (\delta_2/\alpha_0, \infty)$. Moreover, given K_1, a_{12} is determined as the solution of $A_{22} = K_1$, hence, as a two-valued function of K_1, and ${}_I\alpha_1$ $(= \alpha_0 a_{12} + \alpha_1)$ is a linear function of a_{12}. We complete the proof by induction. Assume the theorem to hold for the moments $\{{}_I\alpha_i\}_0^{2n-4}$, that is, for every moment in ${}_I\Delta_{n-1}$. Denote these moments by $\{B_i\}_0^{2n-4}$. Then we may consider ${}_I\Delta_n$ as a function of the $n - 1$ unknowns $a_{1n}, a_{2n}, \cdots, a_{n-1,n}$. To determine these, the conditions of the theorem give

$$\begin{aligned}
&\alpha_0 a_{1n} + \alpha_1 a_{2n} + \cdots + \alpha_{n-2} a_{n-1,n} + \alpha_{n-1} = B_{n-1},\\
&L_2(\alpha_0, \alpha_1)a_{1n} + L_2(\alpha_1, \alpha_2)a_{2n} + \cdots \\
&\qquad + L_2(\alpha_{n-2}, \alpha_{n-1})a_{n-1,n} + L_2(\alpha_{n-1}, \alpha_n) = B_n,\\
&\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\\
&L_{n-2}(\alpha_0, \alpha_1, \cdots, \alpha_{n-3})a_{1n} + L_{n-2}(\alpha_1, \alpha_2, \cdots, \alpha_{n-2})a_{2n} + \cdots \\
&\quad + L_{n-2}(\alpha_{n-2}, \alpha_{n-1}, \cdots, \alpha_{2n-5})a_{n-1,n} + L_{n-2}(\alpha_{n-1}, \alpha_n, \cdots, \alpha_{2n-4}) = B_{2n-4},\\
&L_n(\alpha_0, \alpha_1, \cdots, \alpha_{n-1})a_{1n} + L_n(\alpha_1, \alpha_2, \cdots, \alpha_n)a_{2n} + \cdots \\
&\quad + L_n(\alpha_{n-2}, \alpha_{n-1}, \cdots, \alpha_{2n-3})a_{n-1,n} + L_n(\alpha_{n-1}, \alpha_n, \cdots, \alpha_{2n-2}) = K_{n-1}.
\end{aligned} \tag{24}$$

The first $n - 2$ of these equations may be written as

$$
(25)\qquad
\begin{aligned}
&L_n(\alpha_0, \alpha_1, \cdots, \alpha_{n-1}) &&= B_{n-1},\\
&L_n(\alpha_1, \alpha_2, \cdots, \alpha_n) &&= B_n - a_{12}B_{n-1},\\
&\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots &&\\
&L_n(\alpha_{n-3}, \alpha_{n-2}, \cdots, \alpha_{2n-4}) &&= B_{2n-4} - a_{1,n-2}B_{n-1} - \cdots\\
& && \qquad - a_{n-3,n-2}(B_{2n-3} - \cdots).
\end{aligned}
$$

Denote the constants on the right of (25) by $\{D_i\}_{n-1}^{2n-4}$, and combine with the last equation of (24). We obtain

$$
\begin{aligned}
D_{n-1}a_{1n} + D_n a_{2n} + \cdots + D_{2n-4}a_{n-2,n}&\\
+ (\alpha_{n-2}a_{1n} + \alpha_{n-1}a_{2n} + \cdots + \alpha_{2n-4}a_{n-1,n} + \alpha_{2n-3})a_{n-1,n}&\\
+ (\alpha_{n-1}a_{1n} + \alpha_n a_{2n} + \cdots + \alpha_{2n-3}a_{n-1,n} + \alpha_{2n-2}) = K_{n-1}&.
\end{aligned}
$$

Substituting in this the solution of (25) for $a_{1n}, a_{2n}, \cdots, a_{n-2,n}$ (this solution necessarily exists, since the determinant of the coefficients is $\Delta_{n-2} \neq 0$) in terms of $a_{n-1,n}$ and simplifying, we obtain

$$
(26)\qquad
\begin{aligned}
&\Delta_{n-1}a_{n-1,n}^2 + 2\Delta''_{n-1}a_{n-1,n}\\
&+ \begin{vmatrix} \alpha_0 & \cdots & \alpha_{n-3} & \alpha_{n-1}\\ \cdots & \cdots & \cdots & \cdots\\ \alpha_{n-3} & \cdots & \alpha_{2n-6} & \alpha_{2n-4}\\ \alpha_{n-1} & \cdots & \alpha_{2n-4} & \alpha_{2n-2}\end{vmatrix}
- \begin{vmatrix} \alpha_0 & \cdots & \alpha_{n-3} & D_{n-1}\\ \cdots & \cdots & \cdots & \cdots\\ \alpha_{n-3} & \cdots & \alpha_{2n-6} & D_{2n-4}\\ D_{n-1} & \cdots & D_{2n-4} & 0\end{vmatrix}\\
&\qquad - \Delta_{n-2}K_{n-1} = 0.
\end{aligned}
$$

The condition that the discriminant of this last equation be non-negative gives, when the determinants are combined by means of the Studnička expansion,[9]

$$
(27)\qquad K_{n-1} \geqq \frac{\delta_n}{\delta_{n-1}} - \frac{1}{\delta_{n-2}}\begin{vmatrix} \alpha_0 & \cdots & \alpha_{n-3} & D_{n-1}\\ \cdots & \cdots & \cdots & \cdots\\ \alpha_{n-3} & \cdots & \alpha_{2n-6} & D_{2n-4}\\ D_{n-1} & \cdots & D_{2n-4} & 0\end{vmatrix} \equiv E, \qquad \mathfrak{Z}_n \equiv (E, \infty).
$$

By (26), $a_{n-1,n}$ is determined as a two-valued function of K_{n-1}, and thus we have

$$
\begin{aligned}
\alpha_{2n-3} = L_{n-1}(\alpha_0, \alpha_1, \cdots, \alpha_{n-2})a_{1n} + \cdots + L_{n-1}(\alpha_{n-2}, \alpha_{n-1}, \cdots, \alpha_{2n-4})\, a_{n-1,n}&\\
+ L_{n-1}(\alpha_{n-1}, \alpha_n, \cdots, \alpha_{2n-3})&
\end{aligned}
$$

is a linear function of $a_{1n}, a_{2n}, \cdots, a_{n-1,n}$.

[9] The form of the Studnička expansion here used is shown explicitly in (28′) below. See, for example, E. Pascal, *Die Determinanten*, 1900, pp. 39–40.

In order to derive an important corollary of this theorem, we need

LEMMA 3. If $\Delta_i > 0$ $(i = 1, 2, \cdots)$, *then for* $\{D_i\}$ *arbitrary*

$$
\begin{vmatrix} \alpha_0 & \cdots & \alpha_{n-3} & D_{n-1} \\ \cdots & \cdots & \cdots & \cdots \\ \alpha_{n-3} & \cdots & \alpha_{2n-6} & D_{2n-4} \\ D_{n-1} & \cdots & D_{2n-4} & 0 \end{vmatrix} \leqq 0 \qquad (n = 3, 4, \cdots). \tag{28}
$$

Proof. The lemma obviously holds for $n = 3$. Assume it holds for $n = k - 1$. By the Studnička expansion, we have

$$
\begin{aligned}
&\Delta_{n-3} \begin{vmatrix} \alpha_0 & \cdots & \alpha_{n-3} & D_{n-1} \\ \cdots & \cdots & \cdots & \cdots \\ \alpha_{n-3} & \cdots & \alpha_{2n-6} & D_{2n-4} \\ D_{n-1} & \cdots & D_{2n-4} & 0 \end{vmatrix} \\
&\quad = \Delta_{n-2} \begin{vmatrix} \alpha_0 & \cdots & \alpha_{n-4} & D_{n-1} \\ \cdots & \cdots & \cdots & \cdots \\ \alpha_{n-4} & \cdots & \alpha_{2n-8} & D_{2n-5} \\ D_{n-1} & \cdots & D_{2n-5} & 0 \end{vmatrix} - \begin{vmatrix} \alpha_0 & \cdots & \alpha_{n-4} & D_{n-1} \\ \cdots & \cdots & \cdots & \cdots \\ \alpha_{n-4} & \cdots & \alpha_{2n-8} & D_{2n-5} \\ \alpha_{n-3} & \cdots & \alpha_{2n-7} & D_{2n-4} \end{vmatrix}^2 \leqq 0,
\end{aligned} \tag{28'}
$$

and the lemma follows by induction.

Hence, as a corollary of Theorem IV, we have

COROLLARY. *If* $\Delta_i > 0$ $(i = 1, 2, \cdots)$, *then*

$$
\alpha_{2n} \geqq \frac{\Delta_{n+1}}{\Delta_n} \qquad (n = 1, 2, \cdots). \tag{29}
$$

We are thus led naturally to study the class of all persymmetric polynomials belonging to sets S satisfying the condition $\Delta_n > 0$. Hamburger[10] has shown that $\Delta_i > 0$, for every i, is a necessary and sufficient condition that the moment problem in the interval $(-\infty, \infty)$ have at least one solution with infinitely many points of increase. In this case, as noted above (footnote 8), the moments are defined as $\alpha_i = \int_{-\infty}^{\infty} x^i \, d\psi(x)$, and the associated polynomials are OP. The corollary has thus given, without the introduction of the notion of a weight function, or of an interval of orthogonality, or of mechanical quadratures, that for OP the even moments are positive, and in addition has given us a lower bound for these moments. A further application to OP follows.

The normalizing factors corresponding to any sequence $\{\Phi_n(x)\}$ of OP are[11]

[10] H. Hamburger, *Über eine Erweiterung des Stieltjesschen Momentenproblem*, I, II, III, Math. Annalen, vol. 81(1920), pp. 235–319; vol. 82(1921), pp. 120–164, 168–187. See also M. Riesz, *Sur le probleme des moments*, III, Arkiv f. Mat., Astron., och Fys., vol. 17, no. 16 (1922–23).

[11] J. Shohat, loc. cit.

$a_n = (\delta_n/\delta_{n+1})^{\frac{1}{2}}$. Hence, the set $\{a_n\}$ *is an invariant of any set S made up of OP.* Moreover, (29) gives here

$$\alpha_{2n} \geqq a_n^{-2}; \qquad a_n \geqq \alpha_{2n}^{-\frac{1}{2}} \qquad (n = 1, 2, 3, \cdots). \tag{29'}$$

Thus, for Hermite and Laguerre polynomials we have respectively:[12]

$$a_n \geqq [\Gamma(n + \tfrac{1}{2})]^{-\frac{1}{2}} \quad \text{and} \quad a_n \geqq [\Gamma(2n + \alpha)]^{-\frac{1}{2}}, \qquad (n = 1, 2, 3, \cdots).$$

Returning now to the recurrence relation (7) and the general persymmetric polynomials, as a consequence of (6), we see that *the sequence* $\{\lambda_n\}$ *is an invariant of the set S.* Thus, for any sequence $\{{}_k\Phi_n(x)\}$ of a given set S, (7) may be written as

$$\begin{gathered} {}_k\Phi_n(x) = (x - {}_kc_n)\,{}_k\Phi_{n-1}(x) - \lambda_n\,{}_k\Phi_{n-2}(x) \\ (n = 2, 3, \cdots;\ {}_k\Phi_0 = 1,\ {}_k\Phi_1(x) = x - {}_kc_1), \end{gathered} \tag{30}$$

where λ_n is invariant in S, and only $\{{}_kc_n\}$ varies with $\{{}_k\Phi_n(x)\}$. In particular, the recurrence relation for the symmetric sequence $\{{}_s\Phi_n(x)\}$ of S is

$${}_s\Phi_n(x) = x\,{}_s\Phi_{n-1}(x) - \lambda_n\,{}_s\Phi_{n-2}(x) \qquad (n = 2, 3, \cdots;\ {}_s\Phi_1(x) \equiv x). \tag{31}$$

The recurrence relation (30) leads to another representation of any ${}_k\Phi_n(x)$ of a given set S,[13]

$${}_k\Phi_n(x) = \begin{vmatrix} x - {}_kc_1 & \lambda_2 & 0 & \cdots & 0 & 0 \\ 1 & x - {}_kc_2 & \lambda_3 & \cdots & 0 & 0 \\ 0 & 1 & x - {}_kc_3 & \cdots & 0 & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & \cdots & x - {}_kc_{n-1} & \lambda_n \\ 0 & 0 & 0 & \cdots & 1 & x - {}_kc_n \end{vmatrix}, \tag{32}$$

where the λ_i are invariants in S. If $\lambda_i > 0$ $(i = 2, 3, \cdots)$, dropping the subscript k, we may write (32) in the symmetric form[14] $(n = 1, 2, \cdots)$

$$\Phi_n(x) = \begin{vmatrix} x - c_1 & \lambda_2^{\frac{1}{2}} & 0 & \cdots & 0 & 0 \\ \lambda_2^{\frac{1}{2}} & x - c_2 & \lambda_3^{\frac{1}{2}} & \cdots & 0 & 0 \\ 0 & \lambda_3^{\frac{1}{2}} & x - c_3 & \cdots & 0 & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & \cdots & x - c_{n-1} & \lambda_n^{\frac{1}{2}} \\ 0 & 0 & 0 & \cdots & \lambda_n^{\frac{1}{2}} & x - c_n \end{vmatrix}. \tag{33}$$

[12] The exact values for a_n in the respective cases are $a_n = 2^{\frac{1}{2}n}[\pi^{\frac{1}{2}}\Gamma(n + 1)]^{-\frac{1}{2}}$ and $a_n = [\Gamma(n + 1)\Gamma(n + \alpha)]^{-\frac{1}{2}}$; see J. Shohat, loc. cit., p. 30.

[13] O. Perron, *Die Lehre von den Kettenbrüchen*, 1913, p. 11.

[14] O. Bottema, *Die Nullstellen gewisser durch Rekursionsformeln definierten Polynome*, Akad. Amsterdam, Proc. Sec. Sc., vol. 34(1931), pp. 681–691.

Moreover, $\Delta_i > 0$ $(i = 1, 2, \cdots)$ implies $\lambda_i > 0$ $(i = 2, 3, \cdots)$,[15] and hence, we are again led naturally to consider OP.

We now turn to a study of expression (33) for $\Phi_n(x)$.

2. **The Jacobi determinant expression for orthogonal polynomials; associated orthogonal polynomials.** By definition, a Jacobi matrix $\| a_{ij} \|$ is a matrix in which $a_{ij} = a_{ji}$ and $a_{ij} = 0$ if $i < j - 1$. The representation (33) then leads at once to

THEOREM V. *Any OP $\Phi_n(x)$ is $(-1)^n$ times the characteristic function of the Jacobi matrix $\| a_{ij} \|$, where $a_{ii} = c_i$, and $a_{i,i+1} = -\lambda_{i+1}^{\frac{1}{2}}$.*

Moreover, the matrix of the determinant (33) is itself a Jacobi matrix, and hence, we shall call this representation of $\Phi_n(x)$ the *Jacobi determinant expression* for $\Phi_n(x)$. We shall call the representation (2) for any OP $\Phi_n(x)$ the *persymmetric determinant expression* for $\Phi_n(x)$.

The expression (33) for $\Phi_n(x)$ as a Jacobi determinant is not unique. Since the sets $\{c_i\}_1^n$ and $\{\lambda_i\}_2^n$ constitute a set of $2n - 1$ constants upon which are imposed only the condition $\lambda_i > 0$ $(i = 2, 3, \cdots, n)$ and the n conditions that certain functions of them equal numerically the coefficients of the powers of x in $\Phi_n(x)$, in general there are ∞^{n-1} ways of expressing $\Phi_n(x)$ as a Jacobi determinant of type (33). Thus, for

$$\Phi_2(x) \equiv x^2 + d_1 x + d_0 \equiv \begin{vmatrix} x - c_1 & \lambda_2^{\frac{1}{2}} \\ \lambda_2^{\frac{1}{2}} & x - c_2 \end{vmatrix},$$

c_1, c_2, and λ_2 are determined in terms of d_0 and d_1 by means of the relations: $c_1 + c_2 = -d_1$, $c_1c_2 > d_0$ and $\lambda_2 = c_1c_2 - d_0$.

THEOREM VI. *Given $\delta > 0$ and n, there is a sequence of OP $\{\Phi_n(x)\}$ such that the zeros of $\Phi_n(x)$ differ from any sequence of constants $\{c_i\}_1^n$ by less than δ.*

Proof. Introduce a sequence of OP $\{\tau_n(x)\}$:

$$(34)\quad \tau_n(x) \equiv \begin{vmatrix} x - c_1 & \epsilon^{\frac{1}{2}} & 0 & \cdots & 0 & 0 \\ \epsilon^{\frac{1}{2}} & x - c_2 & \epsilon^{\frac{1}{2}} & \cdots & 0 & 0 \\ 0 & \epsilon^{\frac{1}{2}} & x - c_3 & \cdots & 0 & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & \cdots & x - c_{n-1} & \epsilon^{\frac{1}{2}} \\ 0 & 0 & 0 & \cdots & \epsilon^{\frac{1}{2}} & x - c_n \end{vmatrix}$$

$$\equiv \prod_1^n (x - c_i) + \epsilon P_{n-2}(x) + \epsilon^2 P_{n-4}(x) + \cdots \equiv \pi_n(x) + \epsilon R(x).$$

When ϵ is chosen sufficiently small, the theorem is obvious if we let $\tau_n(x) \equiv \Phi_n(x)$.

[15] If we impose the added condition $\lambda_1 > 0$, then conversely $\lambda_i > 0$ $(i = 1, 2, \cdots)$ implies $\Delta_i > 0$ $(i = 1, 2, \cdots)$.

(35) TABLE OF CONSTANTS c_n $(n \geqq 1)$ AND λ_n $(n \geqq 2)$

Trigonometric polynomials

$$c_n = 0 \qquad \lambda_n = \tfrac{1}{4}\ (n > 2),\ \lambda_2 = \tfrac{1}{2}$$

Legendre polynomials

$$c_n = 0 \qquad \lambda_n = \frac{(n-1)^2}{(2n-1)(2n-3)}$$

Hermite polynomials

$$c_n = 0 \qquad \lambda_n = \tfrac{1}{2}(n-1)$$

Laguerre polynomials

$$c_n = 2n + \alpha - 2 \qquad \lambda_n = (n-1)(n+\alpha-2) \qquad (\alpha > 0)$$

Jacobi polynomials in $(-1, 1)$ with $(\alpha, \beta) > 0$

$$c_n = \frac{(\alpha-\beta)(\alpha+\beta-2)}{(\alpha+\beta+2n-2)(\alpha+\beta+2n-4)}$$

$$\lambda_n = \frac{4(n-1)(n+\alpha+\beta-3)(n+\alpha-2)(n+\beta-2)}{(2n+\alpha+\beta-3)(2n+\alpha+\beta-4)^2(2n+\alpha+\beta-5)}$$

Given any positive integer n, we associate with $\Phi_n(x)$ and any of its representations (33) the system $\{\Theta_i(x)\}_0^n$ of polynomials defined as follows:

$$\Theta_0(x, n) = 1, \qquad \Theta_1(x, n) = x - c_n, \qquad \Theta_2(x, n) = \begin{vmatrix} x - c_{n-1} & \lambda_n^{1/2} \\ \lambda_n^{1/2} & x - c_n \end{vmatrix},$$

(36)
$$\Theta_3(x, n) = \begin{vmatrix} x - c_{n-2} & \lambda_{n-1}^{1/2} & 0 \\ \lambda_{n-1}^{1/2} & x - c_{n-1} & \lambda_n^{1/2} \\ 0 & \lambda_n^{1/2} & x - c_n \end{vmatrix}, \qquad \cdots, \qquad \Theta_n(x, n) \equiv \Phi_n(x),$$

where $\Theta_i(x) \equiv \Theta_i(x, n)$. The $\Theta_i(x)$ satisfy the recurrence relation

(37)
$$\Theta_i(x) = (x - c_i')\Theta_{i-1}(x) - \lambda_i'\Theta_{i-2}(x)$$
$$(i = 2, 3, \cdots, n;\ \Theta_0(x) = 1,\ \Theta_1(x) = x - c_n),$$

where $c_i' = c_{n-i+1}$, $\lambda_i' = \lambda_{n-i+2} > 0$ $(i = 2, 3, \cdots, n)$. The $\Theta_i(x, n)$ are denominators of the successive convergents of the continued fraction

$$\frac{\lambda |}{|x - c_n} - \frac{\lambda_n |}{|x - c_{n-1}} - \frac{\lambda_{n-1} |}{|x - c_{n-2}} - \cdots - \frac{\lambda_2 |}{|x - c_1} \qquad (\lambda > 0 \text{ and arbitrary}).$$

(37) shows that $\{\Theta_i(x)\}_0^n$ represents a finite sequence of orthogonal polynomials, and we have

THEOREM VII. *Corresponding to any representation of a polynomial* $\Phi_n(x)$ *as a Jacobi determinant of type* (33) *there exist two finite systems of orthogonal polynomials* $\{\Phi_i(x)\}_0^n$ *and* $\{\Theta_i(x)\}_0^n$, *where*[16]

$$\Phi_0(x) = 1, \qquad \Phi_1(x) = x - c_1, \qquad \Phi_2(x) = \begin{vmatrix} x - c_1 & \lambda_2^{\frac12} \\ \lambda_2^{\frac12} & x - c_2 \end{vmatrix},$$

$$\Phi_3(x) = \begin{vmatrix} x - c_1 & \lambda_2^{\frac12} & 0 \\ \lambda_2^{\frac12} & x - c_2 & \lambda_3^{\frac12} \\ 0 & \lambda_3^{\frac12} & x - c_3 \end{vmatrix}, \qquad \cdots,$$

$$\Theta_0(x) = 1, \qquad \Theta_1(x) = x - c_n, \qquad \Theta_2(x) = \begin{vmatrix} x - c_{n-1} & \lambda_n^{\frac12} \\ \lambda_n^{\frac12} & x - c_n \end{vmatrix},$$

$$\Theta_3(x) = \begin{vmatrix} x - c_{n-2} & \lambda_{n-1}^{\frac12} & 0 \\ \lambda_{n-1}^{\frac12} & x - c_{n-1} & \lambda_n^{\frac12} \\ 0 & \lambda_n^{\frac12} & x - c_n \end{vmatrix}, \qquad \cdots, \qquad \Theta_n(x) \equiv \Phi_n(x). \tag{38}$$

The sequence of polynomials $\{\Theta_i(x)\}_0^n$ when considered in connection with $\{\Phi_n(x)\}$ shall be called a sequence of *associated orthogonal polynomials*.

For the two sequences $\{\Theta_i(x)\}_0^n$ and $\{\Phi_i(x)\}_0^n$ to coincide it is easily seen that the necessary and sufficient condition is: $c_i = c_{n-i+1}$ $(i = 1, 2, \cdots, n)$ and $\lambda_i = \lambda_{n-i+2}$ $(i = 2, 3, \cdots, n)$. In order that these relations hold for every n, it is necessary that $c_i =$ constant c $(i = 1, 2, \cdots)$ and $\lambda_i =$ constant $\lambda > 0$ $(i = 2, 3, \cdots)$, i.e., that the polynomials satisfy the recurrence relation

$$\Phi_n(x) = (x - c)\Phi_{n-1}(x) - \lambda\Phi_{n-2}(x).$$

The sequence satisfying this relation is

$$\{\lambda^{\frac12 n} (\sin (n + 1) \text{ arc cos } \tfrac12\lambda^{-\frac12}(x - c))/\sin \text{ arc cos } \tfrac12\lambda^{-\frac12}(x - c)\}_0^\infty.$$

In particular, when $c = 0$ and $\lambda = \frac14$, this reduces to the classical sequence

$$\{2^{-n} (\sin (n + 1) \text{ arc cos } x)/\sin \text{ arc cos } x\}_0^\infty.$$

Since a sequence of OP which is determined by a recurrence relation of type (7) forms a Sturm Chain,[17] it follows that

(1) all polynomials (38) have real and distinct zeros;

(2) the zeros of $\Phi_n(x) \equiv \Theta_n(x)$ are separated by the zeros of $\Theta_{n-1}(x)$;

[16] From known properties of the weight functions corresponding to such finite systems (Hamburger, loc. cit.) we conclude the weight functions $\psi_1(x)$ and $\psi_2(x)$ corresponding respectively to $\{\Phi_i(x)\}_0^n$ and $\{\Theta_i(x)\}_0^n$ are step-functions taking exactly $n + 1$ distinct values in the interval $(-\infty, \infty)$. The points of discontinuity of $\psi_1(x)$ and $\psi_2(x)$ coincide with the zeros of $\Phi_n(x)$; and the sum of the saltus is the same for $\psi_1(x)$ and $\psi_2(x)$.

[17] For the properties of a Sturm chain see, for example, J.-A. Serret, *Cours d'Algèbre Supérieure*, Tome I, 1885, pp. 276–305.

(3) the zeros of $\Theta_j(x)$ are separated by the zeros of $\Theta_{j-1}(x)$ $(j = 2, 3, \cdots, n)$;

(4) the zeros of $\Theta_j(x)$ $(j = 1, 2, \cdots, n-1)$ lie within the interval formed by the extreme zeros of $\Phi_n(x)$.

Let $\{x_{ij}\}$ and $\{y_{ij}\}$ $(i = 1, 2, \cdots, j)$ denote respectively the zeros of $\Phi_j(x)$ and $\Theta_j(x)$ arranged in each case in increasing order of magnitude.

Then, considering $\Theta_1(x, n) = x - c_n$, and recalling $x_{1n} < y_{1,n-1} < y_{1,n-2} < \cdots, x_{nn} > y_{n-1,n-1} > y_{n-2,n-2} > \cdots$, we have

$$x_{1n} < c_j < x_{nn} \qquad (j = 1, 2, \cdots, n). \tag{39}$$

Likewise, considering $\Theta_2(x, n) = x^2 - (c_n + c_{n-1})x + c_n c_{n-1} - \lambda_n$, we obtain in a very simple manner

$$\begin{aligned} x_{1n} &< \frac{c_j + c_{j-1} - [(c_j - c_{j-1})^2 + 4\lambda_j]^{\frac{1}{2}}}{2}, \\ x_{nn} &> \frac{c_j + c_{j-1} + [(c_j - c_{j-1})^2 + 4\lambda_j]^{\frac{1}{2}}}{2}, \qquad (j = 2, 3, \cdots, n), \end{aligned} \tag{40}$$

$$x_{1n} < \frac{c_j + c_{j-1}}{2} - \lambda_j^{\frac{1}{2}}, \qquad x_{nn} > \frac{c_j + c_{j-1}}{2} + \lambda_j^{\frac{1}{2}}, \qquad (j = 2, 3, \cdots, n). \tag{41}$$

In particular, if c_i and λ_i, for $i = 1, 2, \cdots, n$ and $i = 2, 3, \cdots, n$, respectively, attain their maxima $c_M = c_M(n)$ and $\lambda_M = \lambda_M(n)$ for the same $j = M$, and their minima c_m and λ_m for the same $j = m$, then (41) gives

$$x_{1n} < c_{m-1} - \lambda_m^{\frac{1}{2}}, \qquad x_{nn} > c_{M-1} + \lambda_M^{\frac{1}{2}}. \tag{42}$$

If $c_n \to c$ and $\lambda_n \to \lambda$ as $n \to \infty$, then since $\lim_{n\to\infty} x_{1n}$ and $\lim_{n\to\infty} x_{nn}$ are known to exist,

$$\lim_{n\to\infty} x_{1n} \leqq c - \lambda^{\frac{1}{2}}, \qquad \lim_{n\to\infty} x_{nn} \geqq c + \lambda^{\frac{1}{2}}. \tag{43}$$

Denote by F_i the i-th convergent of the continued fraction

$$\frac{1 \,|}{|\, \frac{2k-1}{k}} - \frac{1 \,|}{|\, \frac{2k-1}{k}} - \frac{1 \,|}{|\, \frac{2k-1}{k}} - \cdots \qquad (k \geqq 2). \tag{44}$$

Then it may be shown that $F_i < 0$ for some $i \leqq k + 2$.

THEOREM VIII. *Assume* $\lambda_i - \lambda_{n-k} > 0$ *for any* $k < n - 2$ *and for every* $i = n - k + 1, n - k + 2, \cdots, n$, *and either*

Case I. $c_i - c_{n-k-1} > 0$, *or*

Case II. $c_{n-k-1} - c_i > 0$,

for every $i = n - k, n - k + 1, \cdots, n$; *then*

$$x_{nn} > c_{n-k-1} + \frac{2k-1}{k}\lambda_{n-k}^{\frac{1}{2}}, \quad \textit{or} \quad x_{1n} < c_{n-k-1} - \frac{2k-1}{k}\lambda_{n-k}^{\frac{1}{2}}, \tag{45}$$

respectively.

Proof. Consider Case I. In $\Theta_{k+2}(x, n)$ substitute

$$x = c_{n-k-1} + \frac{2k-1}{k}\lambda^{\frac{1}{2}}_{n-k} \equiv v_k .$$

$$\Theta_{k+2}(v_k, n) = | a_{ij} |,$$

where

$$a_{ii} = \frac{2k-1}{k}\lambda^{\frac{1}{2}}_{n-k} - (c_{n-k+i-2} - c_{n-k-1}),$$

$$a_{i,i+1} = a_{i+1,i} = \lambda^{\frac{1}{2}}_{n-k+i-1}, \qquad a_{ij} = 0, \qquad (i - j \neq 0, \pm 1).$$

Subtract $k(2k-1)^{-1}$ times the first column of $| a_{ij} |$ from the second. Then $| a_{ij} |$ can be written as $(2k-1)k^{-1}\lambda^{\frac{1}{2}}_{n-k}$ times a determinant of lower order. Denote the elements of the resulting second column by a'_{2j} $(j = 1, 2, \cdots, k+2)$. Subtract $(a'_{22})^{-1}\lambda^{\frac{1}{2}}_{n-k+1}$ times this column from the third. Then $| a_{ij} |$ can be written as a'_{22} times a determinant of lower order, where

$$a'_{22} \equiv \left(\frac{2k-1}{k} - \frac{k}{2k-1}\right)\lambda^{\frac{1}{2}}_{n-k} - (c_{n-k} - c_{n-k-1})$$

$$< \left(\frac{1\,|}{|\,\frac{2k-1}{k}} - \frac{1\,|}{|\,\frac{2k-1}{k}}\right)^{-1}\lambda^{\frac{1}{2}}_{n-k} \equiv F_2^{-1}\lambda^{\frac{1}{2}}_{n-k}.$$

Similarly, denote the elements of the resulting third column by a'_{3j} $(j = 1, 2, \cdots, k+2)$, and subtract $(a'_{33})^{-1}\lambda^{\frac{1}{2}}_{n-k+2}$ times this third column from the fourth. Then,

$$a'_{33} \equiv \frac{2k-1}{k}\lambda^{\frac{1}{2}}_{n-k} - \lambda_{n-k+1}\left[\left(\frac{1\,|}{|\,\frac{2k-1}{k}} - \frac{1\,|}{|\,\frac{2k-1}{k}}\right)^{-1}\lambda^{\frac{1}{2}}_{n-k} - (c_{n-k} - c_{n-k-1})\right]^{-1}$$

$$- (c_{n-k+1} - c_{n-k-1}) < \left(\frac{1\,|}{|\,\frac{2k-1}{k}} - \frac{1\,|}{|\,\frac{2k-1}{k}} - \frac{1\,|}{|\,\frac{2k-1}{k}}\right)^{-1}\lambda^{\frac{1}{2}}_{n-1} \equiv F_3^{-1}\lambda^{\frac{1}{2}}_{n-k}.$$

In general

$$a'_{ii} < F_i^{-1}\lambda^{\frac{1}{2}}_{n-k} \qquad (i = 2, 3, \cdots, k+2).$$

Since $F_i < 0$ for some $i \leqq k+2$, some element in the sequence a_{11}, a'_{22}, a'_{33}, $\cdots$, $a'_{k+2,k+2}$ must be negative. Denote the first negative element in this sequence by a'_{gg}. Consider the determinant obtainable from $| a_{ij} |$ by striking out its last $k + 2 - g$ rows and columns. This determinant may be written as $\Phi'_g(v_k, n)$, where $\Phi'_g(x, n)$ is a polynomial of the type $\Phi_g(x, n)$ associated with $\Theta_{k+2}(x, n)$ in the manner discussed in Theorem VII. The above method factors $\Phi'_g(v_k, n)$ into $a_{11}a'_{22}a'_{33} \cdots a'_{gg} < 0$. Hence, $\Phi'_g(x, n) = 0$ for some $x > v_k$ and hence also $y_{k+2,k+2} > v_k$. Since $x_{nn} > y_{k+2,k+2}$, the theorem follows for Case I.

By our substituting $x = c_{n-k-1} - (2k - 1)k^{-1}\lambda_{n-k}^{\frac{1}{2}}$ in $\Theta_{k+2}(x, n)$, a similar argument leads, in Case II, to the upper bound for x_{1n}.

COROLLARY. *Let, for $n \to \infty$, $\lambda_n \to \lambda$ and $c_n \to c$, and let every $\lambda_i < \lambda$ and*
Case I. *every $c_i < c$, or*
Case II. *every $c_i > c$;*
then

$$\lim_{n\to\infty} x_{nn} \geqq c + 2\lambda^{\frac{1}{2}} \quad \textit{or} \quad \lim_{n\to\infty} x_{1n} \leqq c - 2\lambda^{\frac{1}{2}}, \tag{46}$$

respectively.

THEOREM IX. *With the notation of* (42),

$$x_{nn} < c_M + 2\lambda_M^{\frac{1}{2}} - \frac{n+1}{2^{n+1}}\lambda_M^{\frac{1}{2}}, \tag{47}$$

$$x_{1n} > c_m - 2\lambda_M^{\frac{1}{2}} + \frac{n+1}{2^{n+1}}\lambda_M^{\frac{1}{2}}. \tag{48}$$

Proof. In (33) substitute $x = c_M + 2\lambda_M^{\frac{1}{2}} - \epsilon \equiv w_n$. We obtain

$$\Phi_n(w_n) \equiv |a_{ij}|,$$

where

$$a_{ii} = c_M - c_i + 2\lambda_M^{\frac{1}{2}} - \epsilon, \qquad a_{i,i+1} = a_{i+1,i} = \lambda_{i+1}^{\frac{1}{2}}, \qquad a_{ij} = 0 \qquad (i - j \neq 0, \pm 1).$$

Subtract $(a_{nn})^{-1}\lambda_n^{\frac{1}{2}}$ times the n-th column of $|a_{ij}|$ from the $(n - 1)$-th. Then we have as a factor of $|a_{ij}|$:

$$c_M - c_n + 2\lambda_M^{\frac{1}{2}} - \epsilon > 0 \qquad \text{if } \epsilon < 2\lambda_M^{\frac{1}{2}}.$$

Denote the elements of the resulting $(n - 1)$-th column by $a'_{n-1,j}$ $(j = 1, 2, \cdots, n)$. Subtract $(a'_{n-1,n-1})^{-1}\lambda_{n-1}^{\frac{1}{2}}$ times the resulting $(n - 1)$-th column of $|a_{ij}|$ from the $(n - 2)$-th. Then as a factor of $|a_{ij}|$ we have

$$a'_{n-1,n-1} \equiv c_M - c_{n-1} + 2\lambda_M^{\frac{1}{2}} - \epsilon - \frac{\lambda_n}{c_M - c_n + 2\lambda_M^{\frac{1}{2}} - \epsilon} > 2\lambda_M^{\frac{1}{2}} - \epsilon$$
$$- \frac{\lambda_M}{2\lambda_M^{\frac{1}{2}} - \epsilon} > \frac{(2\lambda_M^{\frac{1}{2}} - \epsilon)^2 - \lambda_M}{2\lambda_M^{\frac{1}{2}}} > \tfrac{3}{2}\lambda_M^{\frac{1}{2}} - 2\epsilon > 0, \quad \text{if } \epsilon < \tfrac{3}{4}\lambda_M^{\frac{1}{2}}.$$

Continuing this process and using primes to indicate the elements of the altered columns, we obtain

$$|a_{ij}| = a_{nn}a'_{n-1,n-1}a'_{n-2,n-2} \cdots a'_{11}, \tag{49}$$

where $a'_{n-k+1,n-k+1} > (k + 1)k^{-1}\lambda_M^{\frac{1}{2}} - b_k k^{-1}\epsilon > 0$ if $\epsilon < (k + 1)b_k^{-1}\lambda_M^{\frac{1}{2}}$ $(b_k = 2^{k+1} - k - 2)$. Hence, for any $\epsilon < (n + 1)2^{-n-1}\lambda_M^{\frac{1}{2}}$ we have expressed $|a_{ij}|$ as a product of positive factors, and consequently $x_{nn} < w_n$. The lower bound for x_{1n} is obtainable in a similar manner.

COROLLARY. *If $c_M(n) \to c$ and $\lambda_M(n) \to \lambda$, as $n \to \infty$, c and λ finite, then*

$$\lim_{n\to\infty} x_{nn} \leqq c + 2\lambda^{\frac{1}{2}}. \tag{50}$$

Also if $c_m(n) \to c'$ as $n \to \infty$, then

$$\lim_{n\to\infty} x_{1n} \geqq c' - 2\lambda^{\frac{1}{2}}. \tag{51}$$

Combining (46), (50) and (51), we have $\lim_{n\to\infty} x_{nn} = c + 2\lambda^{\frac{1}{2}}$ $(c = \lim_{n\to\infty} c_M(n))$, and $\lim_{n\to\infty} x_{1n} = c' - 2\lambda^{\frac{1}{2}}$ $(c' = \lim_{n\to\infty} c_m(n))$.

Now let $\overline{\lim}_{n\to\infty} c_n = c$ and $\overline{\lim}_{n\to\infty} \lambda_n = \lambda$. Select k such that $c_i < c + \epsilon$ $(i = k, k+1, \cdots)$ and $\lambda_i < \lambda + \epsilon$ $(i = k+1, k+2, \cdots)$. Consider $\Theta_{n-k+1}(x, n)$, n arbitrary. Stieltjes[18] has shown that between any two zeros of an OP $\Phi_j(x)$ there lies at least one zero of $\Phi_{j+r}(x)$ $(r = 1, 2, \cdots)$. Hence, relation (38) implies that at most k zeros of $\Phi_n(x)$ lie outside the interval containing all zeros of $\Theta_{n-k+1}(x, n)$, yielding, since ϵ may be chosen arbitrarily small, $\lim_{n\to\infty} y_{n-k+1,n-k+1}$ $(\leqq c + 2\lambda^{\frac{1}{2}})$ as an upper bound for the interval in which the zeros of $\{\Phi_n(x)\}$ may be everywhere dense. Moreover, as a consequence of Theorem VIII for n, k, and $n - k \to \infty$, the zeros of $\{\Theta_{n-k+1}(x, n)\}_{n=0}^{n=\infty}$ are dense in the neighborhood of $c + 2\lambda^{\frac{1}{2}}$. A similar argument leads to the lower bound, and we have

THEOREM X. *The zeros of $\{\Phi_n(x)\}$ are nowhere dense outside the interval $(c - 2\lambda^{\frac{1}{2}}, c + 2\lambda^{\frac{1}{2}})$, and are dense in the neighborhood of the end points of this interval.*[19]

Expression (33) implies that $\Phi_n(x)$ is the characteristic function of the Jacobi form

$$Q(y, y) = -\sum_{i=1}^{n} c_i y_i^2 + 2 \sum_{i=2}^{n} \lambda_i^{\frac{1}{2}} y_i y_{i-1}. \tag{52}$$

Its spectrum, i.e., the zeros of $\Phi_n(x)$ in (33), has been investigated by Krein.[20]

THEOREM XI (Theorem II of Krein). *If the c_j $(j = 1, 2, \cdots)$ and λ_k $(k = 2, 3, \cdots)$ corresponding to sequences of OP $\Phi_n(x)$ vary independently, then*

$$\frac{\partial x_{in}}{\partial c_k} \geqq 0 \qquad (i, k = 1, 2, \cdots, n). \tag{53}$$

$$\textit{In the sequence } \frac{\partial x_{in}}{\partial \lambda_2}, \frac{\partial x_{in}}{\partial \lambda_3}, \cdots, \frac{\partial x_{in}}{\partial \lambda_n} \qquad (i = 1, 2, \cdots, n) \tag{54}$$

[18] T. J. Stieltjes, *Recherches sur les fractions continues*, Ann. Fac. Sci. Toulouse, vol. 8 (1894), J, pp. 1–122; vol. 9(1895), A, pp. 1–47.

[19] O. Blumenthal (*Ueber die Entwickelung einer willkürlichen Funktion nach den Nennern des Kettenbruches für* $\int_{-\infty}^{0} \frac{\phi(\xi)d\xi}{z - \xi}$, Thesis, Göttingen, 1898) has shown that if $c_n \to c$ and $\lambda_n \to \lambda$ the interval in which the zeros of $\{\Phi_n(x)\}$ are everywhere dense is exactly $(c - 2\lambda^{\frac{1}{2}}, c + 2\lambda^{\frac{1}{2}})$.

[20] M. Krein, loc. cit.

$i - 1$ terms have the sign $+$ and $n - i$ terms the sign $-$, if suitable signs are affixed to possible zero terms.

Corollary (Krein). *x_{1n} is a non-increasing function, and x_{nn} a non-decreasing function, of $\lambda_2, \lambda_3, \cdots, \lambda_n$:*

$$\frac{\partial x_{1n}}{\partial \lambda_i} \leqq 0, \qquad \frac{\partial x_{nn}}{\partial \lambda_i} \geqq 0 \qquad (i = 2, 3, \cdots, n). \tag{55}$$

Theorem XII (Theorem III of Krein).

$$\begin{aligned} c_m - 2\lambda_M^{\frac{1}{2}} \cos \frac{\pi}{n+1} &\leqq x_{1n} \leqq c_M - 2\lambda_m^{\frac{1}{2}} \cos \frac{\pi}{n+1}, \\ c_m + 2\lambda_m^{\frac{1}{2}} \cos \frac{\pi}{n+1} &\leqq x_{nn} \leqq c_M + 2\lambda_M^{\frac{1}{2}} \cos \frac{\pi}{n+1}. \end{aligned} \tag{56}$$

By use of the zeros of the sequence

$$\left\{\lambda^{\frac{1}{2}n} \sin (n+1) \operatorname{arc} \cos \frac{x - c}{2\lambda^{\frac{1}{2}}} \Big/ \sin \operatorname{arc} \cos \frac{x-c}{2\lambda^{\frac{1}{2}}}\right\}_0^\infty$$

this theorem may, of course, be obtained as an immediate corollary of Theorem XI.

Krein's upper bound for x_{nn} and lower bound for x_{1n} are better than the bounds obtained in Theorem IX. The latter theorem was included to indicate more fully how bounds for the zeros of $\Phi_n(x)$ may be obtained from its Jacobi determinant expression by purely algebraic methods. When $c_m(n) \neq c_M(n)$ and $\lambda_m(n) \neq \lambda_M(n)$, it is clear that for a given sequence of OP at most one of Krein's bounds for each zero will be good for n large. But if, for example, $c_M(n) \to c$ and $\lambda_M(n) \to \lambda$, then Krein's bounds together with Theorem VIII, for a suitably chosen k, lead to asymptotic expressions for x_{nn}. Thus, for classical cases we have

1. Hermite polynomials: Choose $k = N((2n)^{\frac{1}{2}})$, where $N(x)$ denotes the next integer $\geqq x$. Then, from (45) and (56),[21]

$$(2n)^{\frac{1}{2}} - \tfrac{3}{2} + O(n^{-\frac{1}{2}}) < {}_Hx_{nn} \equiv -{}_Hx_{1n} < (2n)^{\frac{1}{2}} + (2n)^{-\frac{1}{2}} + O(n^{-1}), \tag{57}$$

i.e., ${}_Hx_{nn} \equiv -{}_Hx_{1n} = (2n)^{\frac{1}{2}} + O(1)$.

2. Laguerre polynomials: Choose $k = N(n^{\frac{1}{2}})$. Then,[21]

$$4n + 2\alpha - 5n^{\frac{1}{2}} + O(1) < {}_Lx_{nn} < 4n + 2\alpha - 5 + O(n^{-1}), \tag{58}$$

i.e., ${}_Lx_{nn} = 4n + 2\alpha + O(n^{\frac{1}{2}})$.

[21] It has been shown that ${}_Hx_{nn} = (2n+1)^{\frac{1}{2}} - 1.8557571(2n+1)^{-1/6} - 0.3443834(2n+1)^{-5/6} - 0.168715(2n+1)^{-3/2} - 0.151965(2n+1)^{-13/6} + O\{(2n+1)^{-17/6}\}$ (F. Zernike, *Eine asymptotische Entwicklung für die grösste Nullstelle der Hermiteschen Polynome*, Amsterdam Academy, Proc. of Sec. Sc., vol. 34(1931), pp. 673–680); and that ${}_Lx_{nn} = 4n + 2\alpha - 3.7115142(4n+2\alpha)^{\frac{1}{3}} + 2.7550676(4n+2\alpha)^{-\frac{1}{3}} + O(n^{-1})$ (V. E. Spencer, *Asymptotic expressions for the zeros of generalized Laguerre polynomials and Weber functions*, this Journal, vol. 3(1937), pp. 667–675).

3. **Properties of the zeros and intervals of orthogonality of sequences of OP in a given set** S. As a consequence of Theorem XI, it follows that if two sequences $\{{}_l\Phi_n(x)\}$ and $\{{}_m\Phi_n(x)\}$ of a set S of OP satisfy the condition ${}_lc_i > {}_mc_i$ ($i = 1, 2, 3, \cdots$), then the zeros of the polynomials of these sequences satisfy the inequality ${}_lx_{in} \geqq {}_mx_{in}$ ($i = 1, 2, \cdots, n$; $n = 1, 2, 3, \cdots$).

This inequality may be used to yield bounds for the zeros of the polynomials of any sequence of a set S of OP in terms of the zeros of the symmetric sequence $\{{}_s\Phi_n(x)\}$ of S. For, if $\{c_i\}_1^n$ correspond to $\{\Phi_i(x)\}_0^n$, in the same set S take $\{{}_M\Phi_i(x)\}_0^n$ and $\{{}_m\Phi_i(x)\}_0^n$, where ${}_Mc_i = c_M$ and ${}_mc_i = c_m$ ($i = 0, 1, 2, \cdots$). These sequences may be obtained from the symmetric sequence $\{{}_s\Phi_i(x)\}_0^n$ of S by replacing x by $x - c_M$ and $x - c_m$, respectively, and hence the zeros $\{x_{in}\}$ of $\{\Phi_n(x)\}$ and the zeros $\{{}_sx_{in}\}$ of $\{{}_s\Phi_n(x)\}$ satisfy the inequality

$$(59) \qquad {}_sx_{in} + c_m \leqq x_{in} \leqq {}_sx_{in} + c_M \qquad (i = 1, 2, \cdots, n).$$

It is known that $\lim_{n\to\infty} x_{1n}$ and $\lim_{n\to\infty} x_{nn}$ both exist, finite or infinite. Define $\lim_{n\to\infty} (x_{nn} - x_{1n})$ as the *interval of orthogonality* of $\{\Phi_n(x)\}$.[22]

THEOREM XIII. *Any given set S of OP contains sequences $\{\Phi_n(x)\}$ whose interval of orthogonality is* $(-\infty, \infty)$.

Proof. Choose $\{c_n\}$ such that $\overline{\lim}_{n\to\infty} c_n = +\infty$ and $\underline{\lim}_{n\to\infty} c_n = -\infty$. This is possible by Theorem III, Corollary 2. By virtue of (39) the corresponding sequence of polynomials is one of the required $\{\Phi_n(x)\}$.

COROLLARY. *If $\{c_n\}$ is unbounded, then the interval of orthogonality of $\{\Phi_n(x)\}$ is infinite.*[23]

THEOREM XIV. *Every set S of OP such that $\{\lambda_n\}$ is bounded contains sequences $\{\Phi_n(x)\}$ for which the length of the corresponding interval of orthogonality (a, b) differs from any preassigned number $l > 0$ by a quantity $d \leqq 2\lambda^{\frac{1}{2}}$, where* $\lambda = \lim_{n\to\infty} \max(\lambda_2, \lambda_3, \cdots, \lambda_n)$.

Proof. Construct $\{c_n\}$ such that it contains two subsequences $\{c_{n_1}\}$ and $\{c_{n_2}\}$ such that $\lim_{n\to\infty} \lambda_{n_1+1} = \lim_{n\to\infty} \lambda_{n_2+1} = \lambda$, $\lim_{n\to\infty} c_{n_1} = \bar{c}$, $\lim_{n\to\infty} c_{n_2} = \underline{c}$, and

$$(60) \qquad \bar{c} \equiv \overline{\lim_{n\to\infty}}\, c_n = l - \lambda^{\frac{1}{2}},$$

$$(61) \qquad \underline{c} \equiv \underline{\lim}_{n\to\infty}\, c_n = \lambda^{\frac{1}{2}}.$$

[22] The above definition is equivalent to the ordinary definition of the *true interval* of orthogonality, i.e., the interval in which the associated weight function must be considered.

[23] See also J. Shohat, *The relation of the classical orthogonal polynomials to the polynomials of Appell*, Amer. Jour. Math., vol. 58(1936), pp. 453-464; Theorem III. This paper appeared after the above results were obtained.

Then combining (41), (56), (60), and (61), we have

$$l \leqq \lim_{n\to\infty} x_{nn} \leqq l + \lambda^{\frac{1}{2}}, \tag{62}$$

$$0 \leqq -\lim_{n\to\infty} x_{1n} \leqq \lambda^{\frac{1}{2}}. \tag{63}$$

But $\lim_{n\to\infty} x_{nn} - \lim_{n\to\infty} x_{1n} = b - a$; and adding (62) and (63), we have

$$l \leqq \lim_{n\to\infty} x_{nn} - \lim_{n\to\infty} x_{1n} \leqq l + 2\lambda^{\frac{1}{2}}.$$

In a sense this also holds for the case that $\{\lambda_n\}$ is unbounded.

THEOREM XV. *If for a set of OP $\{\lambda_n\}$ is unbounded, then the interval of orthogonality corresponding to any sequence $\{\Phi_n(x)\}$ of S is infinite.*

Proof. In (41) let $c_r = \min(c_i, c_{i-1})$, then

$$x_{nn} > c_r + \lambda_i^{\frac{1}{2}} \qquad (i = 2, 3, \cdots, n).$$

Hence, if $\{c_r + \lambda_i^{\frac{1}{2}}\}$ is unbounded the theorem is proved. If, however, $\{c_r + \lambda_i^{\frac{1}{2}}\}$ is bounded, then the negative elements in $\{c_r\}$ must yield an unbounded sequence. But by (39), $x_{1n} < c_r$ $(r = 1, 2, 3, \cdots)$. Hence, $x_{1n} \to -\infty$, and the theorem is also true in this case.

COROLLARY. *If for any sequence $\{\Phi_n(x)\}$ of OP $\{\lambda_n\}$ is unbounded, whether the set $\{c_n\}$ is bounded or not, the interval of orthogonality of $\{\Phi_n(x)\}$ is infinite.*[24]

From Theorems XIII and XIV it will be noted that very little can be said of the interval of orthogonality simply from the condition that $\{\lambda_n\}$ be bounded. However, to satisfy the conditions of both these theorems it was necessary to introduce sequences $\{c_n\}$ for which c_n did not approach a limit. If c_n and λ_n approach limits with $\lim_{n\to\infty} \lambda_n = \lambda$, Theorem X shows that the interval in which the zeros of $\{\Phi_n(x)\}$ can be everywhere dense has a length $4\lambda^{\frac{1}{2}}$. In particular, if every $c_i = 0$, that is, for symmetric polynomials, the interval of density is $(-2\lambda^{\frac{1}{2}}, 2\lambda^{\frac{1}{2}})$.

Blumenthal[25] has shown that if $\lambda_n \to \lambda > 0$ and $c_n \to c$, then the interval $\mathfrak{J}$ within which the zeros of $\{\Phi_n(x)\}$ are everywhere dense is $\mathfrak{J} = (c - 2\lambda^{\frac{1}{2}}, c + 2\lambda^{\frac{1}{2}})$; hence, in any set S of OP for which $\{\lambda_n\}$ has a unique limit point $\lambda > 0$, the interval in which the zeros are everywhere dense is an invariant for all sequences of S for which $\{c_n\}$ has the same unique limit point c.

4. The associated polynomials and continued fractions. It is known that the persymmetric polynomials $\{\Phi_n(x)\}$ are the denominators of the successive convergents $\Omega_n(x)/\Phi_n(x)$ of the continued fraction

$$K(x) \equiv \frac{\lambda_1|}{|x - c_1} - \frac{\lambda_2|}{|x - c_2} - \frac{\lambda_3|}{|x - c_3} - \cdots, \tag{64}$$

[24] Ibid., footnote 23.

[25] O. Blumenthal, loc. cit.

and that the numerators $\{\Omega_n(x)\}$ satisfy the recurrence relation

$$(65)\quad \Omega_n(x) = (x - c_n)\Omega_{n-1}(x) - \lambda_n \Omega_{n-2}(x) \qquad (n = 2, 3, \cdots ; \Omega_0(x) = 0, \Omega_1(x) = \lambda_1),$$

which, except for initial conditions, is the same as (7). Considering the sequence $\{\Omega_n(x)/\lambda_1\}$, we see that the sets S arrange themselves in conjugate pairs S_1 and S_2 such that if $\{\Phi_n(x)\}$ is any sequence of S_1, the corresponding sequence $\{\Omega_n(x)/\lambda_1\}$ lies in S_2.

Restricting ourselves now to OP, let us investigate the finite continued fraction associated with the polynomials $\{\Theta_i(x, n)\}_1^n$. The recurrence relation (37) shows that the $\Theta_i(x, n)$ are denominators of successive convergents of the continued fraction

$$(66)\quad K_n(\Theta(n), x) \equiv \frac{1 |}{| x - c_n} - \frac{\lambda_n |}{| x - c_{n-1}} - \frac{\lambda_{n-1} |}{| x - c_{n-2}} - \cdots - \frac{\lambda_3 |}{| x - c_2} - \frac{\lambda_2 |}{| x - c_1} \equiv \frac{\chi_n(x, n)}{\Theta_n(x, n)}.$$

Moreover, from the theory of continued fractions,

$$(67)\quad \frac{\Phi_{n-1}(x)}{\Phi_n(x)} \equiv \frac{1 |}{| x - c_n} - \frac{\lambda_n |}{| x - c_{n-1}} - \frac{\lambda_{n-1} |}{| x - c_{n-2}} - \cdots - \frac{\lambda_3 |}{| x - c_2} - \frac{\lambda_2 |}{| x - c_1}.$$

Hence, $\chi_n(x, n) \equiv \Phi_{n-1}(x)$.

Sherman[26] has shown that

$$(68)\quad \begin{aligned} K'_n(\Phi, x) &\equiv \frac{\lambda_2 |}{| x - c_2} - \frac{\lambda_3 |}{| x - c_3} - \cdots - \frac{\lambda_{n+1} |}{| x - c_{n+1}} \equiv \frac{P_n(x)}{Q_n(x)}, \\ K_n(\Phi, x) &\equiv \frac{\lambda_1 |}{| x - c_1} - \frac{\lambda_2 |}{| x - c_2} - \cdots - \frac{\lambda_n |}{| x - c_n} \equiv \frac{\Omega_n(x)}{\Phi_n(x)} \end{aligned}$$

imply $\Omega_{n+1}(x) = \lambda_1 Q_n(x)$. Moreover, it is known that the zeros of $\Omega_n(x)$ separate those of $\Phi_n(x)$. Hence, the zeros of the denominator of the l-th convergent of $K'_{n-1}(\Theta(n), x)$ separate the zeros of the denominator of the $(l + 1)$-th convergent of $K_n(\Theta(n), x)$, where $l = 1, 2, \cdots, n - 1$. But $K'_{n-1}(\Theta(n), x) = \lambda_n K_{n-1}(\Theta(n - 1), x)$. Hence, the zeros of $\Theta_{n-i}(x, n)$ are separated by the zeros of $\Theta_{n-1-i}(x, n - 1)$. Proceeding in like manner from $\Theta_{n-1-i}(x, n - 1)$ to $\Theta_{n-2-i}(x, n - 2)$, etc., we have

THEOREM XVI. *In the sequence of $n - i$ polynomials*

$$(69)\quad \Theta_{n-i}(x, n), \Theta_{n-1-i}(x, n - 1), \Theta_{n-2-i}(x, n - 2), \cdots, \Theta_1(x, i + 1) \qquad (i = 0, 1, \cdots, n - 1)$$

[26] J. Sherman, *On the numerators of the convergents of the Stieltjes continued fractions*, Trans. Amer. Math. Soc., vol. 35(1933), pp. 64–87; p. 67.

the zeros of each polynomial are separated by those of the polynomial next succeeding it.

This is a generalization of a known property of OP for the case $i = 0$. Theorem XVI may also be obtained as a direct consequence of the type of argument employed above in deriving (39) and (40), for the polynomials (69) satisfy the recurrence relation

$$(70) \qquad \Theta_{n-i}(x, n) = (x - c_n)\Theta_{n-i-1}(x, n-1) - \lambda_n\Theta_{n-i-2}(x, n-2).$$

This shows further that polynomials (69) are denominators of the successive convergents of the continued fraction

$$(71) \qquad \frac{\lambda_{i+1}}{|x - c_{i+1}} \Big| - \frac{\lambda_{i+2}}{|x - c_{i+2}} \Big| - \cdots - \frac{\lambda_n}{|x - c_n} \Big|.$$

In particular, for $i = 1$, (71) becomes $K'_{n-1}(x)$. Hence, in this case (69) may be written as $\{\Omega_i(x)/\lambda_1\}_1^{n-1}$. That is, for $i = 1$, *sequence* (69), *extended by letting* $\Omega_1 = \lambda_1$, *differs by a constant factor from the sequence of numerators of the first n convergents of* $K(x)$.

A sufficient condition that all sequences of associated polynomials (69) corresponding to a given sequence of OP $\{\Phi_n(x)\}$ be equivalent, in the sense that any two of their polynomials of the same degree be identical, is that $\{\Theta_i(x)\}_1^k \equiv \{\Phi_i(x)\}_1^k$, for $k = 1, 2, \cdots, n$, and hence, that $c_i = c_1$ $(i = 2, 3, \cdots, n)$ and $\lambda_i = \lambda_2$ $(i = 3, 4, \cdots, n)$. In particular, for the symmetric case $(c_i = 0)$, with $\lambda_i = \frac{1}{4}$, as noted above in §2, this requires that

$$\{\Phi_n(x)\} \equiv \{2^{-n} (\sin (n - 1) \operatorname{arc} \cos x)/\sin \operatorname{arc} \cos x\}.$$ [27]

For $i = 2$ sequence (69) is, to a constant factor, the sequence of numerators of the successive convergents of the continued fraction $K'_n(x)$ in (68) in which the denominators of the corresponding convergents give, to a constant factor, the sequence $\{\Omega_n(x)\}$. Proceeding by induction, we obtain the following interpretation for (69):

THEOREM XVII. *If constant factors are disregarded, the polynomials* (69) *for* $i = k, k - 1$ $(k = 2, 3, \cdots, n)$ *are respectively numerators and denominators of successive convergents of continued fractions of type* (64).

Let us now return to the sequence $\{\Theta_i(x, n)\}_0^n$. It bears an interesting relation to the "associated" continued fraction $K(x)$ in (64). The numerators of its successive convergents are given by the relation

$$(72) \qquad \Omega_n(x) = \int_{-\infty}^{\infty} \frac{\Phi_n(x) - \Phi_n(y)}{x - y} \, d\psi(y),$$

where $\psi(y)$ is a bounded, monotone, non-decreasing function in $(-\infty, \infty)$ such that

$$(73) \qquad \int_{-\infty}^{\infty} \frac{d\psi(u)}{x - u} \sim \frac{\lambda_1}{|x - c_1} \Big| - \frac{\lambda_2}{|x - c_2} \Big| - \cdots - \frac{\lambda_n}{|x - c_n} \Big| - \cdots \equiv K(x).$$

[27] See also J. Sherman, ibid., pp. 80–81.

Sherman[28] found

$$\frac{\Phi_n(x) - \Phi_n(y)}{x - y} = L_{n,n-1}(y)\Phi_{n-1}(x) + L_{n,n-2}(y)\Phi_{n-2}(x) + \cdots,$$

$$L_{n,n-1}(y) = 1, \quad L_{n,n-2}(y) = y - c_n, \tag{74}$$

$$L_{n,n-3}(y) = (y - c_{n-1})L_{n,n-2}(y) - \lambda_n L_{n,n-1}(y),$$

$$\cdots\cdots\cdots\cdots\cdots\cdots\cdots$$

Hence,

$$L_{n,n-1}(x) \equiv \Theta_0(x, n), \cdots, L_{n,n-i}(x) \equiv \Theta_{i-1}(x, n), \cdots, L_{n,0}(x) \equiv \Theta_{n-1}(x, n). \tag{75}$$

5. **An extension of Krein's theorem.** From definition (6) of c_m and λ_m it is clear that if the $\{c_i\}_1^{m-1}$ and $\{\lambda_i\}_2^m$ remain fixed while c_m increases, then the induced behavior of $\{\alpha_i\}$ will be: the $\{\alpha_i\}_1^{2m-2}$ remain fixed and α_{2m-1} increases. Similarly, if the $\{c_i\}_1^{m-1}$ and $\{\lambda_i\}_2^{m-1}$ remain fixed while λ_m increases, the induced behavior is: the $\{\alpha_i\}_1^{2m-3}$ remain fixed and α_{2m-2} increases. An increase in α_{2m-1} with the $\{\alpha_i\}_{2m}^{2n-1}$ fixed, however, induces changes in the $\{c_i\}_{m+1}^n$ and $\{\lambda_i\}_{m+1}^n$. Instead, let the $\{\alpha_i\}_{2m}^{2n-1}$ vary with α_{2m-1}. Changes in each $\{\alpha_i\}_{2m}^{2n-1}$ may be uniquely determined from (6), as functions of α_{2m-1}, such that the $\{c_i\}_{m+1}^n$ and $\{\lambda_i\}_{m+1}^n$ remain fixed. A similar statement holds for changes in α_{2m-2}. Defining as *admissible* variations changes in $\{\alpha_i\}_1^{2n-1}$ such that $\{\alpha_i\}_1^{k-1}$ is fixed, α_k increases, and the $\{\alpha_i\}_{k+1}^{2n-1}$ vary in such a way that $\{c_i\}_h^n$ and $\{\lambda_i\}_l^n$ ($h = l = \frac{1}{2}(k+3)$, n odd; $h = \frac{1}{2}(k+2)$, $l = \frac{1}{2}(k+4)$, n even) remain fixed, Theorem II of Krein[29] may be stated in the following manner:

THEOREM XIX. *For any OP and admissible variations in the set* $\{\alpha_i\}_1^{2n-1}$:

$$\frac{\partial x_{in}}{\partial \alpha_k} \geqq 0 \qquad (i = 1, 2, \cdots, n;\ k = 1, 3, \cdots, 2n-1). \tag{76}$$

$$\textit{In the sequence } \frac{\partial x_{in}}{\partial \alpha_2}, \frac{\partial x_{in}}{\partial \alpha_4}, \cdots, \frac{\partial x_{in}}{\partial \alpha_{2n-2}} \qquad (i = 1, 2, \cdots, n) \tag{77}$$

$i - 1$ terms have the sign $+$ and $n - i$ terms the sign $-$, if we affix the suitable signs to possible zero terms.

UNIVERSITY OF PENNSYLVANIA.

[28] J. Sherman, ibid., p. 82.
[29] M. Krein, loc. cit.

THE ALGEBRA OF LATTICE FUNCTIONS

By Morgan Ward

I. Introduction

1. The numerous disconnected results on numerical functions (that is, functions on the positive integers to the complex numbers) which are summarized in the first volume of Dickson's *History* have been welded into a simple and coherent theory by Bell in a series of papers culminating in his *Algebraic Arithmetic* (Bell [1][1]). Bell has shown in detail (see, for example, Bell [2], [3], [4], [5], [6]) that all the various inversion formulas, factorability properties, numerical integrations, and so on, of these functions follow from three basic facts.

I. *The set of all numerical functions form a ring with respect to the operations of addition and Dirichlet multiplication.*

The sum $\sigma = \phi + \psi$ of two numerical functions ϕ and ψ is defined by $\sigma(n) = \phi(n) + \psi(n)$, while their Dirichlet product $\pi = \phi\psi$ is defined by

$$\pi(n) = \sum_{d\delta=n} \phi(d)\psi(\delta). \tag{1.1}$$

II. *The set of all numerical functions ϕ such that $\phi(1) \neq 0$ form a group with respect to Dirichlet multiplication.*

The inverse ϕ^{-1} of ϕ satisfies

$$\sum_{d\delta=n} \phi(d)\phi^{-1}(\delta) = \begin{cases} 1 & \text{if} \quad n = 1; \\ 0 & \text{otherwise.} \end{cases} \tag{1.2}$$

For example, the inverse of the function ζ defined by $\zeta(n) = 1$ for all n is the Möbius function $\mu(n)$.

III. *The set of all factorable functions is closed with respect to the operation of Dirichlet multiplication.*

A function ψ is said to be factorable if

$$\psi(mn) = \psi(m)\psi(n) \qquad \text{if } m, n \text{ are co-prime.} \tag{1.3}$$

It may be shown that the factorable functions form a group with respect to Dirichlet multiplication, on excluding the trivial function ω vanishing for all integers n.

Since the positive integers form a semi-ordered set with respect to the relation x divides y, and indeed a lattice, it is natural to ask whether results of like simplicity and generality hold for functions on semi-ordered sets and lattices. But since both Dirichlet multiplication and factorability depend upon a

Received December 27, 1938.

[1] Numbers in brackets refer to the references at the end of the paper.

multiplicative property of the integers, our way is apparently blocked by the impossibility of introducing a multiplication into an arbitrary lattice.[2]

This difficulty is surmountable by passing to Ore's "quotient structures" (Ore [1]) and the analogous quotient sets of a partially ordered set. For these systems, a multiplication naturally presents itself (Ore [1], p. 426) which enables us to define a "Dirichlet product" of two partially ordered quotient sets, and thereby generalize properties I and II to semi-ordered sets. Our results include Weisner's (Weisner [1]) remarkable Möbius function and his associated inversion formulas for lattices and semi-ordered sets.

For factorable functions over lattices, another type of generalization is possible. For excluding the trivial function zero, we easily see from the fundamental theorem of arithmetic that $f(x)$ is factorable if and only if

$$f(m)f(n) = f([m, n])f((m, n)), \tag{1.4}$$

$$f(1) = 1. \tag{1.41}$$

Here (m, n) and $[m, n]$ denote as usual the greatest common divisor and least common multiple of m and n. Condition (1.4) may be immediately extended to lattices, as the implied multiplications occur in the range of the dependent variable.

We are thus enabled to unify results of Dedekind [1], Birkhoff [1], [2], Glivenko [1], [2] on norms, ranks and distances defined over lattices. In particular, we show that Dedekind's module symbol $(\mathfrak{a}, \mathfrak{b})$ (Dedekind [1], pp. 267–271; Dedekind [2]) and the distance function introduced by Glivenko are definable in terms of one another.

An arithmetical function $f(x)$ is said to be multiplicative if

$$f(mn) = f(m)f(n) \qquad \text{for all integers } m, n. \tag{1.5}$$

By extending this definition to the multiplication of quotient sets defined in §4, we show how we may pass from functions factorable over a set to functions factorable over the quotient set. But the closure property III of factorable functions with respect to Dirichlet multiplication is lost, since we prove that it implies that the lattice must be distributive. Our proof rests upon a useful result in pure lattice theory.

2. An element a of a semi-ordered set or a lattice is said to cover another element c of the set (Birkhoff [1]) if $a \neq c$ and $a \supset x \supset c$ implies $a = x$ or $c = x$. A subset of a semi-ordered set is said to be "complete" if for any two elements a and b of the subset, a covers b in the subset if and only if a covers b in the containing set.

[2] The properties of lattices over which a multiplication exists have been studied by Ward and R. P. Dilworth in some detail (Ward [1], [2]; Dilworth [1], [2], [3]; Ward-Dilworth [1], [2]). But for such simple lattices as the modular and non-modular lattices of order five and Dedekind's free modular lattice of order twenty-eight, no multiplication is definable.

3. We prove

THEOREM 3.1. *Let $\mathfrak{S}$ be a lattice such that in every quotient lattice a/b in which $a \neq b$, there exists an element c covered by a. Then $\mathfrak{S}$ is a modular non-distributive lattice if and only if $\mathfrak{S}$ contains a complete modular sublattice of order five.*

The plan of the paper is sufficiently indicated by the chapter titles. We assume that the reader is familiar with the first part of Ore's fundamental memoir *On the foundation of abstract algebra* (Ore [1]) and also with our previous paper in this Journal (Ward [1]) upon multiplication and residuation in structures. We use the notation and terminology of the latter paper with the substitution of the term "lattice" for the term "structure".

It is a pleasure to acknowledge my indebtedness to many stimulating discussions with Professor E. T. Bell, who first called my attention to the important work of Weisner [1], [2].

II. The ring of functions on semi-ordered sets

4. Let $\mathfrak{S}$ be a semi-ordered set of elements $a, b, \cdots$ with respect to a well-defined ordering relation $x \supset y$, and define equality in $\mathfrak{S}$ as usual by $x = y$ if and only if $x \supset y$ and $y \supset x$. Unequal elements will be called distinct. If $u \neq v$ and if $u \supset x \supset v$ implies $u = x$ or $v = x$, we say u covers v, writing $u > v$, $v < u$. If $\mathfrak{S}$ is ordered, we call $\mathfrak{S}$ a chain.

Given any two elements u, v of $\mathfrak{S}$ such that $u \supset v$, the class $\mathfrak{X}$ of all elements x such that $u \supset x \supset v$ forms a semi-ordered set which we call the *quotient* of v by u. We write $\mathfrak{X} = u/v$, the restriction $u \supset v$ being understood. We make the totality of all quotients $\mathfrak{X}$ of $\mathfrak{S}$ into a semi-ordered set Σ by defining $\mathfrak{X} \supset \mathfrak{Y}$ as follows: If $\mathfrak{X} = u/v$ and $\mathfrak{Y} = z/w$, then $\mathfrak{X} \supset \mathfrak{Y}$ if and only if $u \supset z$ and $v \supset w$. With this ordering, two quotients $\mathfrak{X}$ and $\mathfrak{Y}$ are equal if and only if the classes $\mathfrak{X}$ and $\mathfrak{Y}$ are equal in the set-theoretic sense. The quotients u/u form a partially ordered set isomorphic with $\mathfrak{S}$. We call any such quotient a unit.

Given two quotients $\mathfrak{X} = u/v$ and $\mathfrak{Y} = z/w$ such that $v = z$, we define their product to be the quotient $\mathfrak{P} = u/w$, writing

$$\mathfrak{P} = \mathfrak{X} \cdot \mathfrak{Y}, \quad \text{or} \quad u/w = u/v \cdot v/w.$$

If $v \neq z$, no product is defined. This multiplication is associative but non-commutative save in the trivial case $\mathfrak{X} = \mathfrak{Y} =$ a unit.

If $\mathfrak{S}$ is a lattice, our concepts reduce to the quotient structures introduced by Ore.

5. Now let Γ be an arbitrarily chosen division algebra of characteristic zero, and consider the totality of all well-defined one-valued functions ϕ on Σ to Γ. If $\mathfrak{X} = u/v$, we write

$$\phi\mathfrak{X} = \phi_{uv}$$

for the value of ϕ in Γ which corresponds to $\mathfrak{X}$.

We shall assume from now on

P1. *The number of distinct elements x in every quotient $\mathfrak{X}$ of Σ is finite.*

In particular then, a unit quotient contains precisely one distinct element.

Two set functions ϕ and ψ will be said to be equal if and only if their values ϕ_{uv} and ψ_{uv} are equal in Γ for all $\mathfrak{X} = u/v$ of $\mathfrak{S}$. We write as usual $\phi = \psi$.

We introduce an addition and multiplication for set functions as follows. The sum $\sigma = \phi + \psi$ of two set functions ϕ and ψ is defined by

$$\sigma\mathfrak{X} = \phi\mathfrak{X} + \psi\mathfrak{X}, \qquad \sigma_{uv} = \phi_{uv} + \psi_{uv}.$$

The Dirichlet product $\pi = \phi\psi$ of ϕ and ψ in that order is defined by

$$\pi\mathfrak{X} = \sum_{\mathfrak{U}\cdot\mathfrak{V}=\mathfrak{X}} \phi\mathfrak{U}\psi\mathfrak{V}, \qquad \pi_{uv} = \sum_{u\supset x\supset v} \phi_{ux}\psi_{xv}. \tag{5.1}$$

Here the first summation is extended over all distinct pairs of quotients $\mathfrak{U}$ and $\mathfrak{V}$ whose product is $\mathfrak{X}$, in strict analogy with the Dirichlet multiplication (1.1). In the second summation, the product is taken over all distinct elements x of the quotient set u/v. On occasion, we write $\pi_{uv} = \phi_{ux}\psi_{xv}$, the summation over u/v being indicated by the repeated index x. P1 insures that π is a set function.

The following analogue to I of §1 follows immediately from the definitions.

Theorem 5.1. *The totality of all set functions ϕ on Σ to Γ form a ring.*

The ring has a unit element δ defined by

$$\delta_{uv} = 1 \quad \text{if} \quad u = v; \qquad \delta_{uv} = 0 \quad \text{otherwise},$$

with the characteristic property $\delta\psi = \psi\delta = \psi$. We denote this ring by $\mathfrak{R}$.

6. A function ϕ of the ring $\mathfrak{R}$ is called *proper* if and only if it has an inverse ϕ^{-1} with respect to multiplication, so that

$$\phi\phi^{-1} = \delta \quad \text{or} \quad \phi_{ux}\phi^{-1}_{xv} = \delta_{uv}. \tag{6.1}$$

Theorem 6.1. *A function ϕ is proper if and only if*

$$\phi_{uu} \neq 0 \textit{ for every unit quotient } u/u \textit{ of } \Sigma. \tag{6.2}$$

Proof. Condition (6.2) *is necessary.* For if, for some element a of $\mathfrak{S}$, $\phi_{aa} = 0$, (6.1) gives for $u = v = a$, $0\cdot\phi^{-1}_{aa} = 1$, and this is impossible in a division algebra. *Condition* (6.2) *is sufficient.* For if it is satisfied, then if we put $u = v$ in (6.1), $\phi^{-1}_{uu} = 1/\phi_{uu}$ for all u. Thus $\phi^{-1}\mathfrak{X}$ is defined for all unit quotients u/u. Assume that $\phi^{-1}\mathfrak{X}$ is known for all quotients containing fewer than k distinct elements ($k \geqq 2$), and let $\mathfrak{A} = a/b$ be any quotient containing exactly k distinct elements. Then putting $u = a$, $v = b$ in (6.1) and (6.2), we find that $\phi^{-1}\mathfrak{A} = -1/\phi_{aa} \sum' \phi_{ax}\phi^{-1}_{xb}$, the prime indicating that the term with $x = a$ is to be omitted from the sum. Each quotient x/b contains at most $k - 1$ distinct elements. Hence all the values of ϕ^{-1} in the summation are known so that $\phi^{-1}\mathfrak{A}$ is determined. Hence by induction, $\phi^{-1}\mathfrak{X}$ is known for every $\mathfrak{X}$ of Σ.

Since δ is proper and $(\phi\psi)_{uu} = \phi_{uu}\psi_{uu}$, while Γ contains no divisors of zero, we have immediately the following analogue of II in §1.

THEOREM 6.2. *The set of all proper functions on Σ to Γ forms a group with respect to Dirichlet multiplication.*

7. As illustrations of set functions, consider the following list of functions for any semi-ordered set satisfying condition P1.

TABLE OF SPECIAL FUNCTIONS

	NAME OR SYMBOL	VALUE OF PROPERTY
(i)	zero, ω	$\omega_{uv} = 0$
(ii)	one, δ	$\delta_{uv} = \begin{cases} 0, & u \neq v, \\ 1, & u = v \end{cases}$
(iii)	ζ	$\zeta_{uv} = 1$
(iv)	$\theta = \zeta - \delta$	$\theta_{uv} = \begin{cases} 0, & u = v, \\ 1, & u \neq v \end{cases}$
(v)	Möbius function $\mu = \zeta^{-1}$	$\mu_{ux}\zeta_{xv} = \zeta_{ux}\mu_{xv} = \delta_{uv}$
(vi)	ζ^2	ζ^2_{uv} is the number of distinct elements in the quotient u/v
(vii)	covering function κ	$\kappa_{uv} = \begin{cases} 1, & u = v \text{ or } u > v, \\ 0, & \text{otherwise} \end{cases}$
(viii)	$\lambda = \zeta(\kappa - \delta)$,	λ_{uv} is the number of distinct elements in u/v covered by u, and ν_{uv} is the number covering v
(ix)	$\nu = (\kappa - \delta)\zeta$	

The functions δ, ζ, μ and κ are all proper. It is easily shown that if $\mathfrak{S}$ is a chain, then

$$\mu = (-1)^{\theta}\kappa, \qquad \lambda = \nu = \theta = \tfrac{1}{2}(\mu - \kappa).$$

The characteristic property of the Möbius function is expressed by the relations

$$\phi = \zeta\psi \ (\psi\zeta) \quad \text{implies} \quad \psi = \mu\phi \ (\phi\mu)$$

or more completely (Weisner [1]) as follows:

If $\phi_{uv} = \sum_{u \supset x \supset v} \psi_{xv}$, then $\psi_{uv} = \sum_{u \supset x \supset v} \mu_{ux}\phi_{xv}$.

If $\phi_{uv} = \sum_{u \supset x \supset v} \psi_{ux}$, then $\psi_{uv} = \sum_{u \supset x \supset v} \phi_{ux}\mu_{xv}$.

If we take for $\mathfrak{S}$ the integers 1, 2, 3, $\cdots$, for $x \supset y$ the division relation x divides y, and for Γ the field of complex numbers, and if for any quotient $\mathfrak{X} = u/v$ we always choose $u = 1$, then on writing v for $\mathfrak{X}$, $\mu(n)$ is the Möbius function, $\zeta^2(n)$ is the number of divisors of n, and $\lambda(n)$ is the number of distinct prime factors of n. Another numerical function of importance is the total number of prime factors of n, $\rho(n)$. For $\mathfrak{S}$ a modular lattice with a unit, the generalized function ρ_{uv} is the length of any principal chain joining u and v.

8. If $\mathfrak{S}$ contains only a finite number of distinct elements $u_1, u_2, \cdots, u_k$, the ring $\mathfrak{R}$ may be represented as a matric algebra of order k over Γ. For consider the totality of k-rowed square matrices $\Phi = (\phi_{ij})$ over Γ with the property that $\phi_{ij} = 0$ if $u_i \not\supset u_j$. We can correlate each such Φ with the function ϕ whose values are $\phi_{u_i u_j} = \phi_{ij}$. We write $\Phi \backsim \phi$.

If $\Phi \backsim \phi$ and $\Psi \backsim \psi$, then clearly $\Phi = \Psi$ if and only if $\phi = \psi$, and $\Phi + \Psi \backsim \phi + \psi$. The correspondence also preserves multiplication, for

$$(\Phi\Psi)_{ij} = \sum_{x=1}^{k} \phi_{ix}\psi_{xj}.$$

Every term in this sum in which we do not have both $u_i \supset u_x$ and $u_x \supset u_j$ vanishes. Hence by (5.1),

$$(\Phi\Psi)_{ij} = \begin{cases} (\phi\psi)_{u_i u_j} & \text{if } \; u_i \supset u_j, \\ 0 & \text{otherwise.} \end{cases}$$

This correspondence extends to the group of proper functions, to each of which corresponds a non-singular matrix. In particular, since δ corresponds to the unit matrix (δ_{ij}), we have a method for calculating the inverse of any proper function ϕ by calculating the reciprocal of its matrix Φ.

Consider, for example, the Möbius function over the quotient structure of the modular lattice of order five. If we write for simplicity i for u_i ($i = 1, \cdots, 5$)

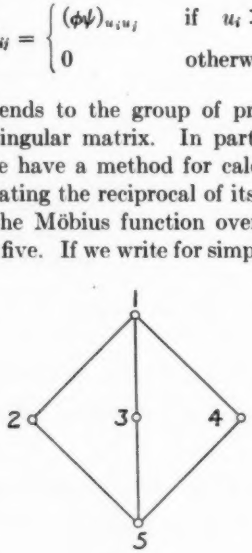

FIG. 1

and designate the elements as in Figure 1, the matrices Z and $M = Z^{-1}$ corresponding to the functions ζ and $\mu = \zeta^{-1}$ are

$$Z = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 \\ 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}, \qquad M = \begin{pmatrix} 1 & -1 & -1 & -1 & 2 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}.$$

Thus $\mu_{11} = 1$, $\mu_{12} = -1, \cdots, \mu_{15} = 2$, and so on.[3]

III. Factorable functions and norms

9. From now on we assume that the semi-ordered set $\mathfrak{S}$ is a lattice and confine ourselves at first to functions on $\mathfrak{S}$ to an Abelian group Δ. For the time being we do not assume postulate P1.

A function ϕ on $\mathfrak{S}$ to Δ is said to be *factorable* if

N1. $a = b$ *in* $\mathfrak{S}$ *implies* $\phi a = \phi b$ *in* Δ;

N2. $\phi a \phi b = \phi(a, b)\phi[a, b]$ *for all pairs of elements* a, b *of* $\mathfrak{S}$.[4]

If we assume that a commutative multiplication xy is definable over the lattice with the properties given in Ward [1], and also assume that

$$\phi e = 1, \qquad e \text{ the unit of } \mathfrak{S}, 1 \text{ the identity element of } \Delta, \tag{9.1}$$

then since $[a, b] = ab$ if $(a, b) = e$ (Ward [1]), we have

$$\phi ab = \phi a \phi b \quad \text{if} \quad (a, b) = e. \tag{9.2}$$

It therefore seems appropriate to call all functions ϕ satisfying N1, N2 "factorable" whether or not a multiplication is definable over $\mathfrak{S}$.

10. Factorable functions are of frequent occurrence in lattice theory. For example, the following functions are always factorable.

(i) The function ζ defined by $\zeta a = 1$ for all a of $\mathfrak{S}$.

(ii) The rank function of Dedekind [1], Birkhoff [1].

(iii) The dimension function of von Neumann [1].

(iv) The norm function of Glivenko.

The last three functions may only be introduced in a modular lattice.

(v) Any evaluation in a residuated lattice is factorable (Ward-Dilworth [3]).

(vi) Let $\mathfrak{S}$ be an Archimedian residuated lattice of order $\geqq 2$. Then $\mathfrak{S}$ contains divisor-free elements, and each element a of $\mathfrak{S}$ has only a finite number λa of divisor-free divisors. λx is (additively) factorable over $\mathfrak{S}$.

[3] By applying the ideas developed in Bell's *Algebraic Arithmetic*, a similar representation by infinite matrices may be given for any denumerable semi-ordered set as no questions of convergence are involved.

[4] The group Δ may be written additively if preferred.

(vii) The ordinary product $\phi\psi$ of two factorable functions defined by $(\phi\psi)x = \phi x\psi x$ is factorable. (The Dirichlet product of factorable functions over a quotient lattice need not be factorable.)

(viii) If $\mathfrak{S}'$ is a sublattice of $\mathfrak{S}$, every function factorable over $\mathfrak{S}$ is factorable over $\mathfrak{S}'$.

(ix) If $\mathfrak{S}$ is a chain, every function satisfying N1 also satisfies N2 and is hence factorable.

(x) Let $\mathfrak{S}$ be a residuated lattice which is the direct product (Ward-Dilworth [2]) of lattices $\mathfrak{S}_\alpha$. Then a function ϕ factorable over $\mathfrak{S}$ defines functions ϕ_α factorable over $\mathfrak{S}_\alpha$. Conversely, the ordinary product of functions ϕ_α factorable over $\mathfrak{S}_\alpha$ gives a function factorable over $\mathfrak{S}$. The instance of common arithmetic occurs when $\mathfrak{S}$ is the direct product of chains. It is essential that $\mathfrak{S}$ be residuated.

11. We have the following fundamental lemma:

LEMMA 11.1. *If ϕ is factorable over $\mathfrak{S}$ and $a \supset b$, then for any c of $\mathfrak{S}$*

$$\phi[a, (b, c)] = \phi(b, [a, c]).$$

For if $a \supset b$, N2, N1 give $\phi a\phi b\phi c = \phi a\phi(b, c)\phi[b, c] = \phi(a, (b, c))\phi[a, (b, c)]\phi[b, c] = \phi(a, c)\phi[b, c]\phi[a, (b, c)]$; $\phi a\phi b\phi c = \phi(a, c)\phi[a, c]\phi b = \phi(a, c)\phi([a, c], b)\phi[[a, c], b] = \phi(a, c)\phi[b, c]\phi(b, [a, c])$. Since Δ is a group, the result follows.

DEFINITION OF A NORM. *A factorable function ϕx on $\mathfrak{S}$ to an Abelian group Δ is said to be a norm if and only if*

N3. *$a \supset b$ in $\mathfrak{S}$ and $\phi a = \phi b$ in Δ imply $a = b$ in $\mathfrak{S}$.*

We denote a norm function by Nx. Lemma 11.1 gives immediately

THEOREM 11.1. *If a norm Nx is definable over $\mathfrak{S}$, then $\mathfrak{S}$ is a modular lattice.*

For $a \supset b$ implies $[a, (b, c)] = (b, [a, c])$.

Since we have trivially $[(a, b), (a, c)] \supset (a, [b, c])$ and $[a, (b, c)] \supset ([a, b], [a, c])$, N3 gives

THEOREM 11.2. *If a norm Nx is definable over $\mathfrak{S}$, then $\mathfrak{S}$ is distributive if and only if*

$$N(a, [b, c]) = N[(a, b), (a, c)] \quad \text{and} \quad N[a, (b, c)] = N([a, b], [a, c]) \tag{11.1}$$

for every set of three elements a, b, c of $\mathfrak{S}$.

The following theorem is also a consequence of N3.

THEOREM 11.3. *If Nx is a norm over $\mathfrak{S}$, then $a \supset b$ if and only if $Na = N(a, b)$ and $Nb = N[a, b]$.*

IV. Modular functions and distance functions

12. Let Nx be a norm on $\mathfrak{S}$ satisfying conditions N1, N2 and N3. The functions Mxy and Dxy defined by

$$(12.1)\qquad Mxy = \frac{Ny}{N(x, y)}, \qquad Dxy = \frac{N[x, y]}{N(x, y)}$$

are called the modular function and distance function associated with the norm Nx. They are connected by the formulas

$$(12.2)\qquad Mxy = D(x, y)y; \qquad Dxy = MxyMyx.$$

The modular function and the distance function have the following properties:[5]

M1. $a = b$ *implies* $Mac = Mbc$ *and* $Mca = Mcb$, *any* c.

M2. $a \supset b$ *if and only if* $Mba = 1$.

M3. $M(a, b)b = Mab$.

M4. $Ma[a, b] = Mab$.

M5. *If* $a \supset b \supset c$, *then* $Mac = MabMbc$.

D1. $a = b$ *implies* $Dac = Dbc$ *and* $Dca = Dcb$, *any* c.

D2. $a = b$ *if and only if* $Dab = 1$.

D3. $Dab = Dba$.

D4. $D(a, b)b = Da[a, b]$.

D5. *If* $a \supset b \supset c$, *then* $Dac = DabDbc$.

These properties are all simple consequences of the properties of a norm. For example, consider M5 and D4. Since $a \supset c$, $(a, c) = a$. Hence $N(a, c) = Na$ by N1. Therefore by (12.1) $Mac = \dfrac{Nc}{Na} = \dfrac{Nb}{Na}\dfrac{Nc}{Nb}$. But since $a \supset b \supset c$, $a = (a, b)$ and $b = (b, c)$. Hence by N1, $Na = N(a, b)$ and $Nb = N(b, c)$. Therefore by (12.1), $Mac = \dfrac{Nb}{N(a, b)}\dfrac{Nc}{N(b, c)} = MabMbc$, and this is M5. For D4, we have $[(a, b), b] = b$, $((a, b), b) = (a, b)$. Hence by (12.1) and N1, $D(a, b)b = \dfrac{Nb}{N(a, b)} = \dfrac{N[a, b]}{Na}$ by N2. But $[[a, b], a] = [a, b]$ and $([a, b], a) = a$. Hence by N1 and (12.1), $Da[a, b] = \dfrac{N[a, b]}{Na}$, or $D(a, b)b = Da[a, b]$.

Let $\mathfrak{S}$ be a lattice with a unit element e. Then if a modular function Mxy is defined on $\mathfrak{S}\mathfrak{S}$ to Δ with Properties M1–M5, it is easy to show that

$$(12.3)\qquad Nx = Mex$$

[5] Properties M2-M5 are given in Dedekind [1]. Properties D2 and D3 are a generalization of the first two distance axioms of Fréchet and Hausdorff.

is a norm. Similarly, if a distance function Dxy is defined on $\mathfrak{S}\mathfrak{S}$ to Δ with Properties D1–D5, then

$$Nx = Dex \tag{12.4}$$

is a norm. Hence we have

THEOREM 12.1. *In a lattice with a unit element, a norm, a distance function, and a modular function are equivalent concepts, each definable in terms of the other.*

The following theorem is also immediate.

THEOREM 12.2. *Let $\mathfrak{S}$ be a lattice with a norm Nx and associated modular functions and distance functions Mxy and Dxy. Then $a \supset b$ if and only if $Mab = Dab$.*

13. In many instances, the group Δ consists of the integers or the real numbers under addition, and the value of the norm is positive or zero. This occurs, for example, in the instances (ii)–(vi) cited in §10. We are thus led to the following additional restrictions on a norm.

A set Λ of elements of the group Δ is said to be an "integral set" if (i) it is closed under group multiplication; (ii) it contains the group identity 1; (iii) for at least one element a of Λ, a^{-1} is not in Λ. We may partially order Δ and hence Λ by the division relation $x \mid y$ where $x \mid y$ if and only if yx^{-1} lies in Λ. (For the case when Λ is the set of positive integers and the group operation is addition, $a \mid b$ if and only if $a \leqq b$.)

A norm Nx on $\mathfrak{S}$ to Δ will be said to be *integral* if and only if

N4. *All the values of Nx lie in a set Λ of integral elements of Δ.*

N5. *$a \supset b$ in $\mathfrak{S}$ implies $Na \mid Nb$ in Λ.*

There exist factorable functions satisfying N1, N2, N4 and N5, but not N3; the simplest example is an evaluation (Ward-Dilworth [3]).

Formulas (12.1) show us that we then have the following conditions on the associated modular function and distance function:

M6-D6. *All the values of Mxy (Dxy) lie in Λ.*

But conversely the norm associated with a modular function (distance function) with Properties M1-M6 (D1-D6) will have Properties N4 and N5. The truth of N4 is obvious from (12.3), (12.4). Consider N5. If $a \supset b$ in $\mathfrak{S}$, then $Mba = 1$. Hence since $e \supset a \supset b$, $Meb = MeaMab$ or $Nb = NaMab$. Since Mab lies in Λ, $Na \mid Nb$ in Λ. The proof for the distance function is similar. Let us call a modular function (distance function) satisfying M1-M6 (D1-D6) "integral". Then Theorem 12.1 becomes

THEOREM 13.1. *In a lattice with a unit element, an integral norm, integral distance function and integral modular function are equivalent concepts, each definable in terms of the other.*

We readily find that for any a, b and c

$$MabMbc = Mac\, \frac{N[b, (a, c)]}{N[(a,b),(b,c)]},$$

$$DabDbc = Dac\, \frac{N[(a, c), b]N([a, b], [b, c])}{N(b, [a, c])N[(a, b), (b, c)]}.$$

Since $[(a, b), (b, c)] \supset [b, (a, c)]$ and $(b, [a, c]) \supset [(a, c), b]$, $[(a, b), (b, c)] \supset ([a, b], [b, c])$, it follows from N5 that

M7. $$Mac \mid MabMbc,$$

D7. $$Dac \mid DabDbc.$$

D7 reduces to the familiar triangle inequality for the distance function when Δ is the set of real numbers $\geqq 0$ and the group operation is addition. M7 is an analogous "triangle inequality" for the modular function.

V. Factorable functions and multiplicative functions over quotient lattices

14. Let Σ be the quotient lattice of $\mathfrak{S}$. A function $\phi\mathfrak{X}$ on Σ to Δ is said to be "factorable" if it satisfies N1 and N2, and "multiplicative" if

$$\phi\mathfrak{X}\cdot\mathfrak{Y} = \phi\mathfrak{X}\phi\mathfrak{Y}. \tag{14.1}$$

Here, as in §4, the product $\mathfrak{X}\cdot\mathfrak{Y}$ of the quotient lattices $\mathfrak{X} = u/v$ and $\mathfrak{Y} = v/w$ is the lattice u/w.

A factorable function $\phi\mathfrak{X}$ on Σ to Δ defines a factorable function on $\mathfrak{S}$ to Δ; for we may define ϕx, $x \in \mathfrak{S}$ to mean ϕ_{xx}, x/x a unit quotient of Σ. We may also pass from a factorable function ϕ on $\mathfrak{S}$ to a factorable function ϕ on Σ. For given any quotient $\mathfrak{A} = a/b$, we define $\phi\mathfrak{A}$ to mean $\dfrac{\phi b}{\phi a}$. $\phi\mathfrak{X}$ is obviously factorable. For if $\mathfrak{A} = a/b$, $\mathfrak{B} = c/d$, $(\mathfrak{A}, \mathfrak{B})$ and $[\mathfrak{A}, \mathfrak{B}]$ are defined as $(a, c)/(b, d)$ and $[a, c]/[b, d]$ (Ore [1]). We call $\phi\mathfrak{X}$ the "extension" of ϕ over Σ.[6]

$\phi\mathfrak{X}$ is multiplicative. For if $\mathfrak{A}\cdot\mathfrak{B} = \mathfrak{C}$, where $\mathfrak{A} = a_1/a_2$, $\mathfrak{B} = b_1/b_2$, $\mathfrak{C} = c_1/c_2$, then since $a_1 = c_1$, $a_2 = b_1$, $b_2 = c_2$, we have by N1 $\phi\mathfrak{C} = \dfrac{\phi c_2}{\phi c_1} = \dfrac{\phi b_2}{\phi a_1}$ $= \dfrac{\phi b_2}{\phi b_1}\dfrac{\phi a_2}{\phi a_1} = \phi\mathfrak{A}\phi\mathfrak{B}$

Conversely, if $\mathfrak{S}$ contains a unit element e and $\phi\mathfrak{X}$ is both factorable and multiplicative over Σ, ϕ is the extension of a factorable function on $\mathfrak{S}$. For given any x in $\mathfrak{S}$, we define ϕx to be $\phi e/x = \phi_{ex}$. Then ϕ is evidently factorable over $\mathfrak{S}$. Since ϕ is multiplicative over Σ, $\phi_{ev} = \phi_{eu}\phi_{uv}$. Hence if $\mathfrak{X} = u/v$, $\phi\mathfrak{X} = \dfrac{\phi v}{\phi u}$.

[6] A simple example is given by the rank function ρ of a modular lattice of finite length (Birkhoff [1]). The extension $\rho\mathfrak{X}$, $\mathfrak{X} = u/v$ is then the length of any principal chain joining u and v.

THEOREM 14.1. *Let $\mathfrak{S}$ be a lattice with a unit element. Then a factorable function $\phi\mathfrak{X}$ on the quotient lattice Σ of $\mathfrak{S}$ to an Abelian group Δ is multiplicative over Σ if and only if it is the extension of a factorable function on $\mathfrak{S}$ to Δ.*

The extension of a norm need not be a norm. For let Nx be a norm on $\mathfrak{S}$, and $N\mathfrak{X}$ its extension over Σ. Let b and c be any two elements of $\mathfrak{S}$ such that $b \not\supset c$. Then $(b, c) \neq b$, $[b, c] \neq c$. However, $\dfrac{Nb}{N(b, c)} = \dfrac{N[b, c]}{Nc}$ by N2. Thus for the two quotients $\mathfrak{A} = (b, c)/b$, $\mathfrak{B} = c/[b, c]$, we have $\mathfrak{A} \supset \mathfrak{B}$, $N\mathfrak{A} = N\mathfrak{B}$, but $\mathfrak{A} \neq \mathfrak{B}$ in contradiction to N3.

VI. Factorable functions and Dirichlet multiplication

15. Let $\psi\mathfrak{X}$ denote from now on a function on the quotient lattice Σ of a lattice $\mathfrak{S}$ satisfying P1 of §5 to the division algebra Γ, and consider the totality of such functions which satisfy the postulates N1 and N2 for factorable functions.

THEOREM 15.1. *If the set of factorable functions ψ on Σ to Γ is closed under Dirichlet multiplication, the lattice Σ must be distributive.*

Proof. We observe first that it suffices to show that the basic lattice is distributive (Ore [1]). Now the function ζ which always equals 1 is obviously factorable. Assume that ζ^2 is also factorable. If $\mathfrak{X} = u/v$, $\zeta^2\mathfrak{X}$ is the number of distinct elements in the lattice $\mathfrak{X}$. By hypothesis then

$$\zeta^2\mathfrak{A}\zeta^2\mathfrak{B} = \zeta^2(\mathfrak{A}, \mathfrak{B})\zeta^2[\mathfrak{A}, \mathfrak{B}] \tag{15.1}$$

for every pair of quotient lattices $\mathfrak{A}$, $\mathfrak{B}$ of Σ.

If $\mathfrak{S}$ is non-modular, $\mathfrak{S}$ contains a non-modular sublattice of order five. On lettering its elements as in Figure 2, we have $a = (b, c) = (b, d)$ and $e = [b, c] = [b, d]$.

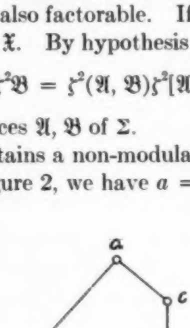

FIG. 2

Consider the three quotient lattices a/b, a/c and a/d. We have then $(a/b, a/c) = (a/b, a/d) = a/a$ and $[a/b, a/c] = [a/c, a/d] = a/e$. We deduce then from (15.1) that

$$\zeta_{ab}^2\zeta_{ac}^2 = \zeta_{aa}^2\zeta_{ae}^2, \qquad \zeta_{ab}^2\zeta_{ad}^2 = \zeta_{aa}^2\zeta_{ae}^2.$$

Hence $\zeta_{ac}^2 = \zeta_{ad}^2$, and this is impossible. Hence $\mathfrak{S}$ is modular.

If $\mathfrak{S}$ is modular but not distributive, then by Theorem 3.1 (see §16), $\mathfrak{S}$ con-

tains a complete modular sublattice of order five. Letter its elements as in Figure 1, with 12345 replaced by *abcde*, respectively. Then we deduce that $\zeta^2_{ab}\zeta^2_{ac} = \zeta^2_{aa}\zeta^2_{ae}$.

But since the sublattice is complete, $\zeta^2_{ab} = \zeta^2_{ac} = 2$, $\zeta^2_{aa} = 1$, $\zeta^2_{ae} \geqq 5$. Hence $\mathfrak{S}$ must be distributive.

16. It remains to prove Theorem 3.1. We may remark that the point of this theorem is that the sublattice is complete; Birkhoff has proved that in any finite modular non-distributive lattice there exists a modular sublattice of order five (Birkhoff [2]). The covering hypothesis of the theorem will be satisfied if the weak ascending chain axiom (Ore [1], p. 410) holds in $\mathfrak{S}$. The theorem may be obviously dualized, but it is easily shown by simple examples that no analogous result is true for the non-modular lattices of order five contained in a non-modular lattice (Dedekind [1]).

Proof of Theorem 3.1. We first show that $\mathfrak{S}$ *contains at least one modular sublattice of order five.* This result, which is purely combinatorial,[7] rests upon the following lemma of Dedekind's (Dedekind [1], p. 252).

Lemma 1. *$\mathfrak{S}$ is distributive if and only if for any three elements a, b and c of*

$$[(a, b), (b, c), (c, a)] = ([a, b], [b, c], [c, a]). \tag{16.1}$$

Now, assume that $\mathfrak{S}$ is modular but not distributive. Then $\mathfrak{S}$ must contain three elements a, b and c such that

$$[(a, b), (b, c), (c, a)] \neq ([a, b], [b, c], [c, a]). \tag{16.2}$$

Let $u = [(a, [b, c]), (b, c)]$, $v = [(b, [c, a]), (c, a)]$, $w = [(c, [a, b]), (a, b)]$. Then

$$\begin{aligned} (u, v) &= (v, w) = (w, u) = [(a, b), (b, c), (c, a)], \\ [u, v] &= [v, w] = [w, u] = ([a, b], [b, c], [c, a]). \end{aligned} \tag{16.3}$$

For consider $[u, v]$. Since $(b, c) \supset (b, [c, a])$ and $(c, a) \supset (a, [b, c])$, $[u, v] = [(a, [b, c]), (b, [c, a])]$. But $(a, [b, c]) \supset [a, c]$ and $b \supset [b, c]$. Hence by the modular axiom, $[u, v] = ([c, a], [b, (a, [b, c])]) = ([c, a], ([b, c], [a, b])) = ([a, b], [b, c], [c, a])$. The remaining equalities in (16.3) follow by symmetry and duality.

No one of the elements u, v and w can divide any other. If for example $u \supset v$, then by the modular axiom $[u, (v, w)] = (v, [u, w])$. Hence by (16.3) $[u, (u, w)] = (v, [v, w])$ or $u = v$. Then by (16.3) $(v, w) = (v, u) = v$, so that $w \supset v$. Hence $u = v = w$, so that (16.3) implies (16.1), contradicting (16.2). The five elements $\{j = (u, v), u, v, w, t = [u, v]\}$ thus form a modular sublattice of $\mathfrak{S}$.

For the second part of the proof, we need the following lemma (Ore [1], p. 419; Birkhoff [1]).

Lemma 2. *If a and b are any two elements of a modular lattice $\mathfrak{S}$, then (a, b) covers a if and only if b covers $[a, b]$.*

[7] The proof of Birkhoff [2] is indirect and rests upon properties of the rank function applicable only if the lattice is of finite length.

With the notation previously employed, the quotient lattices j/u, j/v, j/w and u/t, v/t, w/t are all isomorphic to one another. Hence if $u > t$, the lattice $\{j, u, v, w, t\}$ is complete. If $u \not> t$, then by the hypothesis of the theorem, there exists an element s such that $u > s \supset t$. Let $v' = (v, s)$. Then $(u, v') = j$. By the modular axiom, $[u, v'] = (s, [v, u]) = (s, t) = s$. Then since $u > [u, v']$, $(u, v') > v'$ by Lemma 2. Thus there exists an element v' such that $j > v' \supset v$, $[u, v'] = s$. Similarly, there exists an element w' such that $j > w' \supset w$, $[u, w'] = s$. Let u' be any element such that $j > u' \supset u$. (u' exists by our covering hypothesis.) Then clearly $(u', v') = (v', w') = (w', u') = j$, so that no one of u', v', w' divides any other.

Consider next the three cross-cuts $v'' = [u', v']$, $w'' = [v', w']$ and $[u', w']$. Then if any one divides another, all three are equal. For if, say, $v'' \supset w''$, then since $v' \supset v''$ and $v' > w''$, $v'' = w''$. But then $[[u', v'], w''] = [u', v', w'] = w''$, so that $[u', v'] \supset w''$, $[u', v'] = w''$. Thus $[u', v'] = [v', w'] = [w', u'] = t'$, say, and $\{j, u', v', w', t'\}$ is the desired sublattice.

Assume finally that neither of v'', w'' divides the other. Then $(v'', w'') = u'$, and if we let $t'' = [v'', w''] = [u', v', w']$, then

$$u > s = [u, t''], \qquad u' > v'' > t'', \qquad u' > w'' > t''. \tag{16.4}$$

Let $u'' = (u, t'')$. Then from (16.4) and Lemma 2, $u' \supset u'' > t''$, so that $u'' \neq u'$. By the modular axiom

$$[u'', v''] = [(u, t''), [u', v']] = ([u, [u', v']], t'') = (s, t'') = t'';$$

$$(u'', v'') = ((u, t''), [u', v']) = (u, [u', v']) = [u', (u, v')] = [u', j] = u'.$$

Similarly $[u'', w''] = t''$, $(u'', w'') = u'$. Hence we have shown that there exists a u'' such that $u'' > t''$ and

$$(u'', v'') = (v'', w'') = (w'', u'') = u'; \qquad [u'', v''] = [w'', v''] = [w'', u''] = t''.$$

Consequently $\{u', u'', v'', w'', t''\}$ is the desired sublattice.

References

E. T. Bell.

1. *Algebraic Arithmetic*, New York, 1927.
2. Bull. Am. Math. Soc., vol. 16(1912), pp. 166–167.
3. Bull. Am. Math. Soc., vol. 28(1922), pp. 111–122.
4. Journal Indian Math. Soc., vol. 17(1928), pp. 1–12.
5. Trans. Am. Math. Soc., vol. 25(1924), pp. 145–154.
6. Annali di Matematica, (4), vol. 4(1927), pp. 1–6.

Garrett Birkhoff.

1. Proc. Camb. Phil. Soc., vol. 29(1933), pp. 441–464.
2. Proc. Camb. Phil. Soc., vol. 30(1934), pp. 115–122.

R. P. Dilworth.

1. Bull. Am. Math. Soc., vol. 44(1938), pp. 262–268.
2. Bull. Am. Math. Soc., vol. 44(1938), p. 332, abstract.
3. Bull. Am. Math. Soc., vol. 44(1938), p. 625, abstract.

V. GLIVENKO.
1. Am. Jour. Math., vol. 58(1936), pp. 799-828.
2. Am. Jour. Math., vol. 59(1937), pp. 941-956.
J. VON NEUMANN.
1. Proc. Nat. Acad. Sci., vol. 22(1936), pp. 92-100.
O. ORE.
1. Annals of Math., vol. 36(1935), pp. 406-437.
M. WARD.
1. This Journal, vol. 3(1937), pp. 627-636.
2. Annals of Math., vol. 39(1938), pp. 558-568.
M. WARD AND R. P. DILWORTH.
1. Proc. Nat. Acad. Sci., vol. 24(1938), pp. 162-164.
2. Trans. Am. Math. Soc., vol. 45(1939), pp. 335-354.
3. Annals of Math., vol. 40(1939), pp. 328-338.
L. WEISNER.
1. Trans. Am. Math. Soc., vol. 38(1935), pp. 474-484.
2. Trans. Am. Math. Soc., vol. 38(1935), pp. 485-492.

CALIFORNIA INSTITUTE OF TECHNOLOGY.

MODULAR FIELDS. I

SEPARATING TRANSCENDENCE BASES

BY SAUNDERS MAC LANE

1. **Introduction**. Any extension K of a given field L has a transcendence basis T over L, that is, a set of elements $T = \{t_1, t_2, \cdots\}$ algebraically independent over L and such that all elements of K are algebraic over T. In other words, K can be considered as a (possibly infinite) field of algebraic functions of the variables $t_1, t_2, \cdots$. Many properties of algebraic equations must be restricted to separable equations, without multiple roots, so we enquire: When does a field K have over a subfield L a "separating" transcendence basis T such that all elements of K are roots of *separable* algebraic equations over T?

Forms of this question arise in the analysis of intersection multiplicities for general algebraic manifolds (B. L. van der Waerden [13][1]), in one method of discussing the structure of complete fields with valuations (Hasse and Schmidt [4], p. 16 and p. 46), and in the study of pure forms over function fields (Albert [2]). The properties of such separating transcendence bases may be also considered as one part of a systematic study of the algebraic structure of fields of characteristic p.

A first result, obtained independently by van der Waerden ([13], Lemma 1, p. 620) and by Albert and the author ([2], Theorem 3), is

THEOREM 1. *Any field K obtained by adjoining a finite number of elements to a perfect field P has a separating transcendence basis over P.*

A proof is given in §3 below. The fields treated in this theorem might also be described as finite algebraic function fields of n variables over P, for any integer n. A similar result for a more general ground field is (proof in §7, Theorem 14, Corollary)

THEOREM 2. *If L is a function field of one variable over a perfect coefficient field P, and if K is obtained by adjoining to L a finite number of elements in such a way that every element of K algebraic over L is in L, then K has a separating transcendence basis over L.*

The hypothesis that K is generated over L by a finite number of elements is essential to this theorem. For more general fields K there is a relation between the structure of K and that of its subfields over P.

Received January 13, 1939; presented to the American Mathematical Society, December 29, 1938.

[1] Numbers in brackets refer to the bibliography at the end of the paper.

THEOREM 3. *If a field K has a finite separating transcendence basis over a perfect subfield P, then any intermediate field M, such that $K \supset M \supset P$, also has a separating transcendence basis over P.*

The proof is given in §3.

Our chief purpose is to obtain further theorems of this type for an extension K/L in which the base field L is not restricted to be a perfect field P. Thus we obtain in Theorem 13 of §7 necessary and sufficient conditions that a given subset T be a separating transcendence basis for a given extension K/L, and also necessary and sufficient conditions that a given extension K/L have some finite separating transcendence basis (Theorem 14 in §7).

Extensions K/L with separating transcendence bases are special instances of extensions K/L which "preserve p-independence". The notion of the p-independence of subsets of L, due to Teichmüller ([10], §3), is formulated in §4. Those extensions K/L which "preserve" this p-independence, in the sense that p-independent subsets of L remain p-independent in K, can be characterized in a large number of different ways (Theorem 7 in §4, Theorem 10 of §5, Theorem 16 of §7). They arise naturally here and also in the study of the relative structure of discrete complete fields with valuations (Mac Lane [6]). Such extensions K/L which preserve p-independence seem to have most of the properties appertaining to arbitrary extensions of perfect base fields. An instance is the highly useful possibility of reducing certain equations involving only p-th powers of elements of K (see Theorem 10 of §5).

Extensions preserving p-independence provide a natural tool for investigating separating transcendence bases, largely because, given an extension K/L which preserves p-independence, any intermediate extension M/L with $M \subset K$ must automatically preserve p-independence. This analysis yields in §8 a theorem on separating transcendence bases for intermediate fields, which requires no mention of p-independence in its formulation, and which states that Theorem 3 above is valid without the hypothesis that the base field P be perfect.

In terms of p-independence we introduce for any extension K/L certain "relative p-bases" closely related to separating transcendence bases. In §9 we show that the notion of a p-basis can be used to generalize a theorem of A. A. Albert on pure forms, and we give another property of p-bases.

The earliest result on separating transcendence bases is one of F. K. Schmidt's [8] who proved that a function field of one variable over a perfect field P has a separating transcendence basis over P. His proof (§3 below) gives also the slightly more general statement:

THEOREM 4. *If a field K has transcendence degree 1 over its maximal perfect subfield P, then K has a separating transcendence basis over P.*

It would be tempting to conjecture that this theorem remains true without the restriction that the transcendence degree of K/P is 1. That this conjecture is false we show by a somewhat involved example given in §10. The difficulty of the example resides chiefly in the problem of explicitly computing the maximal

perfect subfield of a field specifically given. The method used in this example (and also two corollaries to Theorem 19 of §8) throws some light on this general problem. Our example also provides an instance of a field whose "degree of imperfection", in the sense of Teichmüller, differs from its transcendence degree over its maximal perfect subfield.

2. **Preliminaries**. **Notation**. All fields to be considered will have characteristic a fixed prime p. If K is such a field, K^{p^n} denotes the field of all p^n-th powers of elements from K. Similarly, if T is any set of elements, T^p is the set of all p-th powers of elements of T, etc. If a field K and certain sets $S, T, \cdots$ are all contained in a larger field, then $K(S, T, \cdots)$ denotes the field obtained by adjoining to K all elements of S, of T, etc. If K contains a subfield L, properties of K relative to the ground field L will be called properties of the "extension" K/L. $K \cap M$ denotes the intersection of the fields K and M, $S \cup T$ the union of the sets S and T, $\{t\}$ the set whose only element is t.

DEFINITIONS.[2] Let K/L be a given extension. An element b of K is *separable* over L if b satisfies a polynomial equation $f(x) = 0$, with coefficients in L, having no multiple roots. If b in K has some power b^{p^e} in L, then b is *purely inseparable* over L. The extension K/L is *separable* (or, *purely inseparable*) if every b in K is separable (or, purely inseparable) over L. A set $T \subset K$ is *algebraically independent* over L if no element t of T is algebraic over the field $L(T - \{t\})$, where $T - \{t\}$ denotes the set T with t deleted. A *transcendence basis* T for K/L is a subset of K algebraically independent over L such that K is algebraic over $L(T)$. A *separating transcendence basis* for K/L is a transcendence basis T such that K is separable over $L(T)$. It can be shown (Mac Lane [5], Theorem 2.1) that K/L has a separating transcendence basis if and only if the elements of K can be so well ordered that each element is either transcendental or separable over the field obtained by adjoining to L all prior elements in the order. The *transcendence degree* for K/L is the (cardinal) number of elements in any transcendence basis for K/L.

A polynomial $f(x_1, \cdots, x_m)$ with coefficients in K may involve some variable x_1 only as powers of x_1^p; the polynomial f is then called *inseparable* in x_1. There is a largest integer e such that f can be written as a polynomial $g(x_1^{p^e}, x_2, \cdots, x_m)$ involving x_1 only as a p^e-th power; this largest p^e is called the *exponent* of x_1 in f. (If x_1 fails to appear in f, use $p^e = \infty$.) If b is inseparable over a field L, the irreducible equation $f(x) = 0$ for b over L is known to be inseparable (exponent $p^e > 1$) in x.

A field P is perfect if $P^p = P$, that is, if each element of P has a unique p-th root in P. The *maximal perfect subfield* of a given (imperfect) field K is the intersection K^{p^∞} of all the fields K^{p^n} ($n = 1, 2, 3, \cdots$).

A known result is (Teichmüller [10], Theorem 12; Mac Lane [5], §2)

LEMMA 1. *An element b which is both separable and purely inseparable over a field L lies in that field.*

[2] Cf. Albert [1], Chapter 7, or B. L. van der Waerden [12], Chapter 5.

3. **Bases over perfect ground fields.** Schmidt's Theorem for function fields of one variable over a perfect base field may be stated in a form including Theorem 4 of the introduction thus:

THEOREM 5. *If an imperfect field K has transcendence degree 1 over a perfect subfield P, then K has a separating transcendence basis over P.*

Proof. Let t be a transcendence basis for K/P. If every root $t^{p^{-e}}$ were in K, K would be algebraic over the perfect field $P(t, t^{p^{-1}}, t^{p^{-2}}, \cdots)$, so K itself would be perfect,[3] counter to assumption. Let e be the largest integer for which $u = t^{p^{-e}}$ is in K. Then $u^{1/p}$ is not in K, and u is a transcendence basis for K/P. It is also a separating transcendence basis for K/P unless K contains an x inseparable over $P(u)$. For such an x the irreducible equation $g(x^p, u) = 0$ over P involves only p-th powers of x, hence can be written as the p-th power of a polynomial over $P(u^{1/p})$. Thus the adjunction of $u^{1/p}$ to $P(u)$ reduces the degree of x over $P(u)$ by a factor p. This means that $u^{1/p}$ must be in $P(u, x) \subset K$, contrary to the choice of e.

A typical necessary and sufficient condition for the existence of a separating transcendence basis is

THEOREM 6. *A field K with a finite transcendence basis T over a perfect subfield P has a separating transcendence basis over P if and only if there is an integer e for which K^{p^e} is separable over $P(T)$.*

Proof. If K does have a separating transcendence basis X over P, each element x of this basis has some power $x^{p^{e'}}$ separable over $P(T)$. X, like T, has but a finite number[4] of elements, so we can choose e as the largest such exponent e' and have X^{p^e} separable over $P(T)$. As K is separable over $P(X)$, K^{p^e} is separable over $P(X^{p^e})$ and hence over $P(T)$, by the transitivity of separability.[5] The necessity of our condition is thereby established.

Conversely, let e be the smallest integer such that K^{p^e} is separable over $P(T)$. If $e = 0$, T itself is a separating transcendence basis, and we are done. Assume then that $e > 0$. We propose to replace $T = \{t_1, \cdots, t_n\}$ by a separably "equivalent" basis $\{u_1^p, u_2, \cdots, u_n\}$ in which one element is a p-th power u_1^p; for then the basis $u_1, \cdots, u_n$ with u_1 replacing u_1^p will "separate" more of K.

Since e was chosen as small as possible, there is in K an element y such that y^{p^e} but not $y^{p^{e-1}}$ is separable over $P(T)$. The irreducible equation $f(y, T) = 0$ for y over $P(T)$ then involves y only in the form y^{p^e}; i.e., has exponent p^e in y. We can also assume that $f(y, T)$ is irreducible in the ring of all polynomials in y and T with coefficients in P. All these coefficients are p-th powers in the perfect field P. If every variable t of T had exponent p or greater in f, f would contain only p-th powers, so could be written as a p-th power of another polynomial, in contradiction to its assumed irreducibility. If t is one of the variables

[3] Steinitz [9], §13, no. 4.

[4] The number of elements in a transcendence basis of K/P is an invariant of K/P. See Steinitz [9], §23.

[5] Steinitz [9], §13, no. 9.

of T which actually appears in f with exponent 1, we can consider $f(y, T) \equiv f(y, t, T - \{t\}) = 0$ as an equation for t over $P(T, y^{p^e})$. According to the Gauss Lemma, this equation is irreducible, while it has exponent 1 in t, hence is separable. Therefore t is separable[6] over $P(T - \{t\}, y^{p^e})$.

The set $T_1 = (T - \{t\}) \cup \{y\}$ obtained from T by replacing t by y is thus another transcendence basis for K/P, such that every element separable over $P(T)$ is separable over $P(T_1)$, by the transitivity of separability (see footnote 5). Furthermore y is not separable over $P(T)$ but is separable over $P(T_1)$. The subfield K_{T_1} of all elements b of K separable over $P(T_1)$ is thus larger than the subfield K_T of all b separable over $P(T)$. Because T is finite, the degree of K over this original subfield K_T is finite,[7] so a repetition of this transition from T to T_1 will finally yield a basis T_m over which all of K is separable.

Theorem 1 of the introduction is an immediate corollary of this theorem. Theorem 3 concerning intermediate fields is also a corollary. For let $K \supset L \supset P$, where K has a finite separating transcendence basis over P. Pick a transcendence basis X for L/P and a similar basis Y for K/L. Then the union $T = X \cup Y$ is a transcendence basis for K/P. The necessary condition of Theorem 6 applied to this basis T then shows that the sufficient condition of Theorem 6 must be satisfied by the original basis X. Hence the intermediate field L does in fact have a separating transcendence basis over P.

4. **Extensions preserving p-independence**. With any given extension K/L there is related in an invariant fashion a purely inseparable extension $K/L(K^p)$. The latter may be analyzed by the concepts of p-independence and p-basis introduced by Teichmüller in §3 of [10]. A subset X of K is *relatively p-independent* in K/L if $K^p(L, X')$ is a proper subfield of $K^p(L, X)$ whenever X' is a proper subset of X. A *relative p-basis* B for K/L is a relatively p-independent set such that $K = K^p(L, B)$. This notion of p-independence has the usual properties of an abstract dependence relation.[8] Therefore, every extension K/L has a relative p-basis and any two relative p-bases for the same extension have the same (cardinal) number of elements. Furthermore any set relatively p-independent in K/L can be embedded in a p-basis for K/L.

A subset X of K is (absolutely) *p-independent* if $X' < X$ implies $K^p(X') < K^p(X)$, where $<$ denotes proper inclusion. An (absolute) *p-basis* B for K is a p-independent subset for which $K = K^p(B)$. This is the special case of the above "relative" definitions obtained by assuming L perfect, for then $L = L^p$, $L \subset K^p$, and the field $K^p(L, X)$ used above becomes[9] $K^p(X)$. Ex-

[6] The argument underlying this exchange—from y^{p^e} separable over $P(T - \{t\}, t)$ to t separable over $P(T - \{t\}, y^{p^e})$—can be stated generally; cf. Lemmas I and II in Mac Lane [5].

[7] By Steinitz [9], Theorem 3 in §13 one proves $K_T \subset K \subset K_T(T^{p^{-e}})$, where $K_T(T^{p^{-e}})$ is certainly an extension of finite degree over K_T.

[8] Van der Waerden [12], p. 204, or Mac Lane [7], §6. The results stated above are given by Theorem 12 of [7] applied to the extension $K/K^p(L)$.

[9] Teichmüller's paper [10] considered chiefly this special case.

tensions which preserve the (absolute) p-independence of subsets of L will now be characterized in several different ways.

THEOREM 7. *Any two of the following properties of an extension K/L are equivalent*:

(i) *Every set $X \subset L$ p-independent in L is p-independent in K.*

(ii) *There is a p-basis B of L which is p-independent in K.*

(iii) $L^p(S) = L \cap K^p(S)$ *for every finite subset* $S \subset L$.

(iv) $L'(L^p) = L \cap L'(K^p)$ *for every subfield* $L' \subset L$.

Condition (iv) on the intersection $L \cap L'(K^p)$ states in effect that the adjunction to L' of p-th powers of elements of K yields no elements in L not obtainable by adjoining p-th powers from L.

DEFINITION. Any extension K/L with one of the equivalent properties (i), (ii), (iii), or (iv) will be said to *preserve p-independence*.

Proof. That (i) implies (ii) is trivial, so consider (ii) → (i). Were some p-independent subset X of L not p-independent in K, there would be an element x in X contained in $K^p(X')$, where X' is a finite subset of X not containing x. Since B of (ii) is a p-basis of L, there is a finite subset $B_1 \subset B$ such that x and X' are in $L^p(B_1)$. This means that $X' \cup \{x\}$ is p-dependent on B_1. In such circumstances one can exchange the elements x and $x' \in X'$ successively with suitable elements[10] of B_1, until all of B_1 is p-dependent on X', x and a remaining subset $B_2 \subset B_1$. This means that

$$L^p(B_1) = L^p(X', x, B_2) \tag{1}$$

and that the combined set $X' \cup \{x\} \cup B_2$ is p-independent in L. The two finite sets B_1 and $X' \cup \{x\} \cup B_2$ are mutually p-dependent over L and have the same number of elements, by construction. But the subset B_1 of B is assumed to remain p-independent in K, while the other subset $X' \cup \{x\} \cup B_2$ is p-dependent over K because we supposed x to be in $K^p(X')$. This is a contradiction since this makes $K^p(B_1)$ have a degree p^{β_1} over K^p, if β_1 is the number of elements in B_1, while the equal field $K^p(X', x, B_2) = K^p(X', B_2)$ has a smaller degree over K^p. Therefore (ii) → (i) in our theorem.

To demonstrate (iii) → (i), suppose counter to (i) that some p-independent set X of L becomes p-dependent in K, so that again some $x \in K^p(X')$. Therefore x is in the intersection $L \cap K^p(X') = L^p(X')$, by (iii). This result states that x is p-dependent on X' over L, counter to assumption.

The implication (iv) → (iii) may be obtained trivially by setting $L' = P(S)$, where P is used to denote some perfect subfield of L.

Finally, to prove that (i) → (iv), let L' be any subfield of L and pick a relative p-basis T for $L'(L^p)/L^p$. Then $L'(L^p) = L^p(T)$. If the conclusion $L \cap L'(K^p)$

[10] By the "exchange" property for dependence relations: If x depends on C and d, but not on C alone, then d depends on C and x. Cf. Teichmüller [10], §3 or Mac Lane [7], §6 and Axiom (E_1).

$= L'(L^p)$ of (iv) were false, there would be an element y of L in $L'(K^p) = K^p(L') = K^p(T)$ but not in $L'(L^p) = L^p(T)$. Thus y is not p-dependent on the set T in L, while T is by construction p-independent in L, so that the usual properties of dependence make[11] $T \cup \{y\}$ a p-independent subset of L. Condition (i) then insures that $T \cup \{y\}$ is p-independent in K, in conflict with the previous assertion that y is in $K^p(T)$. Hence (i) $\to$ (iv). The various implications (ii) $\leftrightarrow$ (i) $\to$ (iv) $\to$ (iii) $\to$ (i) completely establish the theorem.

Such an independence-preserving extension might be viewed as a generalization of the ordinary separable algebraic extensions, in the following sense:

THEOREM 8. *If K is an algebraic extension of L, then K/L preserves p-independence if and only if K/L is separable.*

Proof. If K/L is separable, each p-basis of L remains a p-basis of K (see footnote 14), and this insures that p-independence is preserved. Conversely, suppose that K/L preserves p-independence but is not separable, and denote by K_s the subfield consisting of those elements of K which are separable over L. Then $K \supset K_s$ and K/K_s is purely inseparable,[12] so K contains a c not in K_s with c^p in K_s. Any p-basis B of L is also a p-basis of K_s, so c^p is p-dependent on B, thus lies in $K_s^p(B)$. The exchange property (see footnote 10) of p-dependence provides for an exchange of c^p with some $b \in B$, with the result that $b \in K_s^p(B - \{b\}, c^p) \subset K^p(B - \{b\})$. This states that the set B is not p-independent in K, in violation of the assumption that the extension K/L preserves the p-independence of B. Hence the theorem is proved.

Other examples of extensions which preserve p-independence will now be cited (cf. Mac Lane [6], §6), but for our purposes it is especially important to note that any extension of a perfect field always preserves p-independence.

THEOREM 9. (a) *An extension K/L preserves p-independence if L is perfect or if K/L has a separating transcendence basis.*

(b) *If L is an extension of transcendence degree 1 over a perfect field P, then an extension K/L preserves p-independence if and only if no element of K is inseparable and algebraic over L. In particular, K/L preserves p-independence if L is relatively algebraically closed[13] in K.*

Proof. We first prove (a). If L is perfect, each p-basis of L is void, hence necessarily remains p-independent in any K. If K/L has a separating transcendence basis and if B is a p-basis of L, then B is part of a p-basis[14] of K, hence does remain p-independent in the extension K/L.

Part (b) refers especially to fields L which are function fields of one variable

[11] Mac Lane [7], corollary to Theorem 2.

[12] Steinitz [9], §14, Theorem 1.

[13] A subfield L of K is relatively algebraically closed in K if every element of K algebraic over L lies in L.

[14] By the following theorems of Teichmüller [10], §3: *If K is a separable algebraic extension of L, any p-basis of L is a p-basis of K. If $L(T)$ is a purely transcendental extension of L by algebraically independent elements T, then $B \cup T$ is a p-basis of $L(T)$ if B is a p-basis of L.* Both can be proved readily from the appropriate definitions.

over a perfect coefficient field P. Suppose first that K contains no element inseparable over L. If L were perfect, K/L would preserve p-independence by part (a), so we can assume that L is imperfect. By Theorem 5, L/P then has a separating transcendence basis of one element, t. This element is also (see footnote 14) a p-basis of L, so that if K/L were not to preserve p-independence, t would be in K^p. Then $t^{1/p}$ is in K, but is inseparable over L, contrary to the assumed character of L. Hence K/L preserves p-independence.

Conversely, suppose that K/L does preserve p-independence but that some b in K is inseparable over L. We can again suppose L imperfect and t a separating transcendence basis for L/P. The irreducible equation $f(y, t) = 0$ for $y = b$ over $P[t]$ then has an exponent $p^e > 1$ in y, but because of its irreducibility has exponent 1 in t. Viewed as an equation[15] for t over $P[y]$, it shows that t is separable over $P(b^{p^e}) \subset K^p$. But t is also purely inseparable over K^p, hence t is in K^p, in conflict with the hypothesis that K/L preserves the p-independence of the p-basis $\{t\}$. This completes the proof of part (b) of the theorem.

5. **Equations involving p-th powers**. The essential intrinsic property of extensions K/L which preserve p-independence is the possibility of reducing those algebraic equations between elements of K which involve only coefficients from L and p-th powers of elements from K. This includes a known, simple property of perfect fields L.

THEOREM 10. *A necessary and sufficient condition that an extension K/L preserve p-independence is that, for every finite subset Y of K, the linear dependence over L of the set Y^p implies the linear dependence over L of the set Y itself.*

Proof. Suppose first that K/L preserves p-independence, and that the elements $y_1, \cdots, y_m$ of some set Y have their p-th powers linearly dependent over L. Select a p-basis B of L, so that $L = L^p(B)$, and choose a finite subset $U \subset B$ with the property that $y_1^p, \cdots, y_m^p$ are linearly dependent over $L^p(U)$. If this is true for U the null set, $y_1, \cdots, y_m$ are certainly linearly dependent over L. Otherwise we can successively delete elements from U till we find a new U' and an element u such that $y_1^p, \cdots, y_m^p$ are linearly dependent over $L^p(U', u)$ but not over $L^p(U')$. Since u has degree p over $L^p(U')$, the given linear dependence relation may be expressed as

$$\sum_{i=1}^{m} \left(\sum_{j=0}^{p-1} b_{ij} u^j \right) y_i^p = 0, \qquad b_{ij} \in L^p(U'),$$

where not all $b_{ij} = 0$. Therefore

$$\sum_{j=0}^{p-1} \left(\sum_{i=1}^{m} b_{ij} y_i^p \right) u^j = 0. \tag{1}$$

If one of the coefficients $\sum b_{ij} y_i^p$ in this equation is not 0, (1) is a separable equation for u over $K^p(U')$, so that u is p-dependent on U' in K, in violation

[15] Compare the "exchange" argument for Theorem 6.

of the hypothesis that K/L preserves p-independence. If the coefficients of all powers u^j in (1) are zero, any one of these coefficients which involves a $b_{ij} \neq 0$ provides a linear dependence between $y_1^p, \cdots, y_m^p$ over $L^p(U')$, in contradiction to the choice of u. Hence $y_1, \cdots, y_m$ are linearly dependent over L.

Conversely, suppose that the linear dependence of a set Y^p always implies that of Y, but that K/L does not preserve p-independence. Then any p-basis B of L must become p-dependent in K, so that some b is in $K^p(b_1, \cdots, b_n)$, where $b, b_1, \cdots, b_n$ are distinct elements of B. The algebraic extension $K^p(b_1, \cdots, b_n)$ has over K^p a linear basis consisting of all power products $c = b_1^{e_1} \cdots b_n^{e_n}$ with exponents $e_k = 0, 1, \cdots, p-1$. If $c_1, \cdots, c_m$ are all such power products, $b \in K^p(b_1, \cdots, b_n)$ means that there are elements y_i not all zero in K for which

$$b + y_1^p c_1 + \cdots + y_m^p c_m = 0. \tag{2}$$

Among the elements $1, y_1^p, \cdots, y_m^p$ of K^p pick a linearly independent basis, over L^p, consisting of $1, z_1^p, \cdots, z_t^p$, so that each y_i^p may be written as

$$y_i^p = d_{i0}^p + d_{i1}^p z_1^p + \cdots + d_{it}^p z_t^p, \qquad d_{ij} \text{ in } L.$$

If these expressions are substituted in (2) and the coefficients of each z_i^p collected, one finds

$$\left(b + \sum_{i=1}^m d_{i0}^p c_i\right) + \left(\sum_{i=1}^m d_{i1}^p c_i\right) z_1^p + \cdots + \left(\sum_{i=1}^m d_{it}^p c_i\right) z_t^p = 0. \tag{3}$$

Here the first term $b + \sum d_{i0}^p c_i$ cannot be zero, because b is not p-dependent on $b_1, \cdots, b_n$ over the original field L. Therefore (3) asserts that $1, z_1^p, \cdots, z_t^p$ are linearly dependent over L. The hypothesis of the theorem then shows that $1, z_1, \cdots, z_t$ are linearly dependent over L. It therefore follows that $1, z_1^p, \cdots, z_t^p$ are linearly dependent over L^p and the construction of the z^p's as linearly independent over L^p is contradicted. This implies that the p-basis B of L remains p-independent, which is to say that the given extension does preserve p-independence.

Lemma 2. *Let the extension K/L preserve p-independence, and let quantities $t_1, \cdots, t_n$ be algebraically independent over L, while t_0 in K is algebraically dependent on $t_1, \cdots, t_n$ according to a relation $f(t_0, t_1, \cdots, t_n) = 0$, with coefficients in L. If f is irreducible over L as a polynomial in the variables $t_0, \cdots, t_n$, then f necessarily has exponent 1 in at least one of these variables.*

Proof. The conclusion states in effect that such an irreducible f cannot be a polynomial in the p-th powers $t_0^p, \cdots, t_n^p$. If this were the case, f would be a linear relation between the p-th powers of a certain set of distinct power-products y_i

$$f(t_0, \cdots, t_n) \equiv \sum_{i=1}^m a_i y_i^p = 0, \qquad y_i = t_0^{e_{0i}} \cdots t_n^{e_{ni}}, \qquad (i = 1, \cdots, m),$$

where all the coefficients $a_i \neq 0$. This linear dependence of $y_1^p, \cdots, y_m^p$ over L implies by Theorem 10 a linear dependence of $y_1, \cdots, y_m$:

$$g(t_0, \cdots, t_n) \equiv \sum_{i=1}^{m} b_i y_i = 0, \qquad b_i \text{ in } L.$$

Not all $b_i = 0$, so some t_j, say t_n, must actually appear in g. The degree d of f in this quantity is then at least p times the degree of g in t_n. By the Gauss Lemma d is the degree of the element t_n of K over the field $L(t_0, \cdots, t_{n-1})$, although $g = 0$ provides an equation of smaller degree for t_n over that field. With this contradiction to the assumption that f contained only p-th powers the lemma is established.

6. **Relative p-bases.** Further preliminaries are necessary for the subsequent exposition of the close connection between relative p-bases of an extension K/L in the sense of §4 and separating transcendence bases for the same extension. A first result is the algebraic independence of p-bases, which was established by Teichmüller ([10], Theorem 15) for the case of absolute p-bases.

THEOREM 11. *The elements of any relative p-basis B of an extension K/L which preserves p-independence are algebraically independent over L.*

Proof. Were the elements of B algebraically dependent, one could find elements $t_0, t_1, \cdots, t_n$ in B algebraically dependent but with $t_1, \cdots, t_n$ algebraically independent over L. An irreducible polynomial relation $f(t_0, \cdots, t_n) = 0$ between these quantities must then as in Lemma 2 contain one variable, say t_n, of exponent 1. This equation provides an irreducible and separable equation for t_n over the field $L(t_0, \cdots, t_{n-1}) \subset K^p(L, t_0, \cdots, t_{n-1})$. Over the latter field $K^p(L, t_0, \cdots, t_{n-1})$, $t_n = (t_n^p)^{1/p}$ is also purely inseparable, so that t_n must lie in this field (Lemma 1, §2). This conclusion makes t_n relatively p-dependent on $t_0, \cdots, t_{n-1}$, contrary to the hypothesis on $B \supset \{t_0, \cdots, t_n\}$.

Explicit relative p-bases can be found from absolute p-bases in specific cases by the following process of composition and decomposition.

THEOREM 12. *If an extension K/L preserves p-independence and if B and C are disjoint subsets of K with $C \subset L$, then any two of the following statements imply the third:*

(i) *B is a relative p-basis of K/L;*

(ii) *C is a p-basis for*[16] *L;*

(iii) *$B \cup C$ is a p-basis for*[16] *K.*

Proof. We prove first that (i) & (ii) $\rightarrow$ (iii). The union $B \cup C$ might be p-dependent in two ways. In the first place, an element b of B might lie in $K^p(B - \{b\}, C) \subset K^p(L, B - \{b\})$, but this would violate the relative p-inde-

[16] From this theorem it is also possible to obtain a similar but more general theorem in which statements (ii) and (iii) concern relative p-bases for L/M and K/M respectively, where $K \supset L \supset M$ and the extension L/M is assumed to preserve p-independence (Theorem 12 is the case when M is perfect).

pendence of the set B. In the second place, an element c of C might lie in $K^p(B, C - \{c\})$. There then are distinct elements $b_1, \cdots, b_m$ in B such that

$$c \in K^p(b_1, \cdots, b_m, C - \{c\}) \tag{1}$$

and such that the statement (1) would be false were any b_i omitted. At least one b_i is present in (1) ($m \geqq 1$) because C is known to be p-independent in L and therefore in K. Over the smaller field $K^p(b_2, \cdots, b_m, C - \{c\})$ both c and b_1 have the degree p, so (1) yields an "exchanged" statement

$$b_1 \in K^p(c, b_2, \cdots, b_m, C - \{c\}) = K^p(b_2, \cdots, b_m, C).$$

This type of p-dependence has already been led to a contradiction, so $B \cup C$ is in fact p-independent in K. That it forms a p-basis for K is then readily shown.

The converse implication (ii) & (iii) → (i) is trivial, granted the hypothesis $B \cap C = 0$. As for the third implication, (i) & (iii) → (ii), the p-independence of the set C in L results at once from its assumed p-independence (hypothesis (iii)) in the larger field K. Were C not a p-basis of L, there would be in L an x not in $L^p(C)$. By (iii), x is in $K^p(B, C)$, so there are distinct elements $b_1, \cdots, b_m$ in B such that

$$x \in K^p(b_1, \cdots, b_m, C), \tag{2}$$

but such that the statement (2) would be false were any b_i omitted. If $m > 0$, one deduces as in the previous argument from (1) that $b_1 \in K^p(x, b_2, \cdots, b_m, C) \subset K^p(L, b_2, \cdots, b_m)$, a violation of the relative p-independence of B in K/L (hypothesis (i)). Thus (2) is $x \in K^p(C)$, although we had assumed $x \in L^p(C)$ false. This states that the p-independent set (see footnote 11) $C \cup \{x\}$ of L has become p-dependent over K, contrary to the basic assumption on this extension K/L. The theorem is thereby completely proved.

We can now describe particular relative p-bases in two typical extended fields (involving transcendental, separable, and inseparable adjunctions).

Lemma 3. *A separating transcendence basis S for an extension K/L is always a relative p-basis for K/L.*

For, any p-basis B of L gives rise to a p-basis (see footnote 14) $B \cup S$ for K and this, by the decomposition Theorem 12, makes S a relative p-basis for K/L.

Lemma 4. *If $K \supset K_0 \supset L$, where K/K_0 is a finite purely inseparable extension and K/L preserves p-independence, then the (cardinal) number of elements in a relative p-basis for K/L is the same as the number of elements*[17] *in a relative p-basis for K_0/L.*

Proof. It suffices to consider the case when the degree $[K : K_0]$ is p, so that $K = K_0(a)$, where a^p is in K_0. We wish to construct for K/L a relative p-basis

[17] For the absolute case (L perfect) this lemma has been proved by M. Becker [3].

containing this element a; to this end we show first that a^p is relatively p-independent in K_0/L. Otherwise a^p would be in $K_0^p(L) = K_0^p(B)$, where B is a p-basis for L. Since a^p is not in K_0^p, a^p can here be exchanged with an element b of B, which means that $b \in K_0^p(a^p, B - \{b\}) \subset K^p(B - \{b\})$, contrary to the p-independence of B in L and K.

Since a^p is relatively p-independent in K_0/L, there is a relative p-basis C_0 for K_0/L containing a^p. The replacement of a^p by a in the set C_0 yields, as may be verified from the definitions, a p-basis C for K/L. C and C_0 have the same number of elements, so our lemma is proved.

7. **Criteria for separating transcendence bases.** The notions of p-independence will now be applied to obtain two types of theorems: first, necessary and sufficient conditions that a given set T be a separating transcendence basis for a given extension K/L; secondly, necessary and sufficient conditions that there exist some separating transcendence basis for a given extension.

THEOREM 13. *If the extension K/L preserves p-independence, then a subset $T \subset K$ is a separating transcendence basis for K/L if and only if T is both a transcendence basis for K/L and relatively p-independent in K/L.*

Proof. Lemma 3 insures that any separating transcendence basis has the two specified properties. Conversely, suppose that some T with these two properties is not a separating transcendence basis for K/L. Then some b of K is not separable over $L(T)$, hence satisfies for $y = b$ an irreducible polynomial equation $f(y, T) = 0$ with exponent at least p in y and with coefficients in L. At least one variable t of T must (by Lemma 2 of §5) have exponent 1 in this polynomial f. As in the "exchange" argument for Theorem 6, we can regard $f(y, T) = 0$ as an irreducible and separable equation for t over $L(T - \{t\}, b^p)$, for f involves $y = b$ only as y^p. Therefore t is separable over the larger field $K^p(L, T - \{t\})$. But t is also purely inseparable over this field, so, by Lemma 1, t must lie in the field $K^p(L, T - \{t\})$. This conclusion states that t is relatively p-dependent on $T - \{t\}$, contrary to the hypothesis that T is relatively p-independent. Hence T must be a separating basis.

THEOREM 14. *If the field K has a finite transcendence basis T over its subfield L, then K has a separating transcendence basis over L if and only if*

(i) *K/L preserves p-independence;*

(ii) *for some integer e, $L(K^{p^e})$ is separable over $L(T)$.*

Proof. That condition (i) is necessary was established in Theorem 9(a), while the necessity of (ii) results from the finiteness of the set T exactly as in Theorem 6 of §2.

Conversely, suppose that (i) and (ii) hold, and pick a relative p-basis B for K/L. According to (i) and Theorem 11 the elements of B are algebraically independent over L, so that we can[18] embed B in a transcendence basis $B \cup X$

[18] In any (abstract) dependence relation, an independent subset can be enlarged to form a maximal independent subset. Mac Lane [7], Theorem 3.

for K/L. From the assumed separability of $L(K^{p^e})$ over the original transcendence basis T one finds, since T is finite, a larger integer f such that $L(K^{p^f})$ is separable over $L(B, X)$. Thence we deduce successively the separability of each of the following extensions

$$(1)\quad L^p(K^{p^{f+1}})/L^p(B^p, X^p);\quad L(K^{p^{f+1}})/L(B^p, X^p);\quad K^{p^{f+1}}(L, B)/L(B, X^p).$$

But B is a relative p-basis for K/L, hence $K = K^p(L, B) = K^{p^2}(L, B) = \cdots = K^{p^{f+1}}(L, B)$. Thus (1) states that K is separable over the basis $B \cup X^p$, which is obviously a transcendence basis because $B \cup X$ is by construction a transcendence basis. Thus we have found, as required, a separating transcendence basis[19] $B \cup X^p$ for K/L.

Corollary. *If L is a field of transcendence degree 1 over a perfect subfield P, and if K is an extension of L of finite transcendence degree, containing no elements inseparable and algebraic over L, then K has a separating transcendence basis over L.*

This follows at once from Theorems 14 and 9(b); it includes Theorem 2 of the introduction as the special case when L is a function field of one variable over P. It is also possible to prove this corollary without using the notion of p-independence, by a suitable extension of the exchange process used for Theorem 6. The hypothesis that L is a function field of only one variable is essential to this theorem; for suppose instead that $L = P(x, y)$ is a rational function field of two independent variables x and y over a perfect field P, and consider the extension $K = L(z, u)$, where z is transcendent over L and u is a root of the inseparable equation $u^p = y + xz^p$. Any separating transcendence basis for K/L would consist of a single element t. Let $f(u, t) = 0$ and $g(z, t) = 0$ be respectively the separable irreducible polynomial equations for u and z over $L[t]$, of respective exponents p^α and p^β in t. Then u is separable over $L(t^{p^\alpha})$, z is separable over $L(t^{p^\beta})$, and t^{p^β} is separable over $L(z)$. If $\alpha \geqq \beta$, u is separable over $L(z)$, although the given equation $u^p = y + xz^p$ is inseparable. A similar contradiction arises if $\beta \geqq \alpha$. Hence this extension K of a function field L of two variables can have no separating transcendence basis t.

The close and natural relation between extensions preserving independence and extensions with separating transcendence bases will now be documented with a pair of theorems, one of which gives a necessary and sufficient condition for the existence of a separating basis, while the other gives conditions for the preservation of p-independence.

Theorem 15. *Let K be a field obtained by adjoining a finite number of elements to L. Then K/L preserves p-independence if and only if K has a separating transcendence basis over L. Furthermore, if K/L does preserve p-independence, then a subset T of K is a separating transcendence basis for K/L if and only if it is a relative p-basis for K/L.*

[19] B alone is a transcendence basis, for X must be void. Any x in X is in K, hence by (1) is separable over $L(B, X^p)$, although x manifestly satisfies the inseparable equation $z^p - x^p = 0$ over $L(B, X^p)$, a contradiction to the assumption that X is not void.

Proof. We know that the presence of a separating transcendence basis S for K/L makes K/L preserve p-independence (Theorem 9(a)) and makes S a relative p-basis (Lemma 3 in §6). Thus we need only consider a relative p-basis T for an extension K/L which does preserve p-independence, and prove that T is a separating transcendence basis. If X is any transcendence basis for K/L, and if K_s is the subfield of all elements of K separable over $L(X)$, then X is a separating transcendence basis for K_s/L, hence a relative p-basis for K_s/L (Lemma 3 in §6). But K must be a purely inseparable extension of K_s, so that K/L has by Lemma 4 of §6 a relative p-basis Y consisting of exactly m elements, where m is the number of elements in X. The two p-bases Y and T both have the same number of elements,[20] so that we have in T a set of elements algebraically independent (Theorem 11) and equal in number to the number m of elements in a transcendence basis X. Therefore T is also a transcendence basis for K/L, so must be a separating basis by the sufficient condition of Theorem 13. Theorem 15 is established.

COROLLARY. *If an extension K/L has a finite separating transcendence basis, then any relative p-basis for K/L is a separating transcendence basis for K/L.*

Example. The hypothesis that the transcendence degree of K/L is finite is necessary for the validity of this corollary, even if we restrict the base field L to be perfect, as may be seen by the following example. Let P be a perfect field over which the elements $t_0, x_1, x_2, x_3, \cdots$ are algebraically independent, and introduce the quantities t_n successively as the roots of the (inseparable) equations

$$t_n^p = t_{n-1} + x_n \qquad (n = 1, 2, 3, \cdots).$$

K is the field $P(x_1, x_2, \cdots, t_0, t_1, t_2, \cdots)$ generated by all these quantities, and $T = \{t_0, t_1, t_2, \cdots\}$ is a separating transcendence basis for K/P. Furthermore, the set $X = \{x_1, x_2, \cdots\}$ can be shown to be a p-basis for K, hence a relative p-basis for K/P. Nevertheless, this p-basis is not even a transcendence basis, so is certainly not a separating transcendence basis.

THEOREM 16. *An extension K/L preserves p-independence if and only if $L(y_1, \cdots, y_n)$ has a separating transcendence basis over L for every finite set of elements $y_1, \cdots, y_n$ from K.*

Proof. If K/L does preserve p-independence, the extension from L to the subfield $K' = L(y_1, \cdots, y_n)$ must also preserve p-independence, so that Theorem 15 at once gives a separating transcendence basis for K'/L. Conversely, suppose that each K'/L has such a basis, but that the whole extension K/L does not preserve p-independence, so that some p-independent subset X of L becomes p-dependent in K. This means that some $x \in K^p(X')$, where $X' \subset X$ is a finite subset not containing x. Therefore $x \in L^p(y_1^p, \cdots, y_n^p, X')$ for suitable elements $y_1, \cdots, y_n$ in K. This result states that the original set X has

[20] This number is an invariant of the extension K/L. Mac Lane [7], Theorem 6.

already become p-dependent in the field $K' = L(y_1, \cdots, y_n)$, contrary to the hypothesis that every subfield with such a finite generation has a separating transcendence basis and hence preserves p-independence.

The distinction between extensions preserving p-independence and extensions with separating transcendence bases arises only for extensions with infinitely many generators, as for instance in the extension $K = L(x^{p^{-1}}, x^{p^{-2}}, \cdots)$ (where x is transcendental over L), which preserves p-independence but which, according to Theorem 14, does not have over L a separating transcendence basis.

8. **Intermediate fields.** The question next to be considered is this: If an extension K/M has a separating transcendence basis, and if L is a field between K and M, under what circumstances does L have a separating transcendence basis over M? Our answer, though dependent on the previous analyses of p-independence, can be stated independently of that notion.

THEOREM 17. *If the fields $K \supset L \supset M$ are such that K has a finite separating transcendence basis over M, then L also has a (finite) separating transcendence basis over M.*

Proof. Certainly K/M and thus L/M preserves p-independence (Theorem 9(a)). Pick transcendence bases X and Y respectively for K/L and L/M; $X \cup Y$ is then a transcendence basis for K/M. By the necessary condition (Theorem 14) for the existence of a separating basis, $M(K^{p^e})$ is separable over $M(X, Y)$ for some e. Therefore $M(L^{p^e})$ is separable over $M(X, Y)$. The adjunction of the indeterminates X cannot reduce any equations irreducible over $M(Y)$, so $M(L^{p^e})$ is also separable over $M(Y)$. This is the sufficient condition of Theorem 14 for the existence of a separating transcendence basis for L/M.

Example. This conclusion could not be asserted were the transcendence degree of K/M infinite, even if we restrict M to be a perfect field P. For let $T = \{t_0, t_1, \cdots\}$ be a set of elements algebraically independent over P, and define another set $Y = \{y_2, y_3, \cdots\}$ successively by the inseparable equations

$$y_n^p = t_{n-2} + t_{n-1}t_n^p \qquad (n = 2, 3, 4, \cdots). \tag{1}$$

Then the field $L = P(T, Y)$ does not have a separating transcendence basis[21] over P. Nevertheless L can be embedded in a larger field $K = L(T^{1/p}) = P(T, Y, T^{1/p}) = P(T^{1/p})$ which does have the separating transcendence basis $T^{1/p}$ over P. This counter-example depends essentially on the fact that the intermediate field L also has an infinite transcendence degree.

COROLLARY. *If the fields $K \supset L \supset M$ are such that K has a separating transcendence basis S over M and L has a finite transcendence degree over M, then L has a separating transcendence basis over M.*

[21] Proof given in Mac Lane [5], Lemma 8.5, where the present L appears as a field S_1, and where it is shown that L has a separating transcendence basis neither over P nor over $P(t_0)$.

Proof. Since L/M has a finite transcendence basis, all elements of this basis are algebraic over a finite subset S_0 of S. Hence L is contained in the field K_0 composed of those elements of K algebraic over $M(S_0)$. For this field, S_0 is a finite separating transcendence basis, so Theorem 17 applies to $K_0 \supset L \supset M$.

The question of a separating basis for K itself over the intermediate field L may now be considered.

THEOREM 18. *If $K \supset L \supset M$, where K has a finite separating transcendence basis over M, K has a separating transcendence basis over L if and only if K/L preserves p-independence.*

Proof. The necessity of this condition is known (Theorem 9(a)). Conversely, suppose that K/L does preserve p-independence, that T is any transcendence basis for K/L, and that S is the given separating basis for K/M. Because S is finite, there is an integer e such that S^{p^e} is separable over $L(T)$. Then $L(K^{p^e})$ is separable over $L(T)$, so that K/L must have a separating transcendence basis by the fundamental criterion of Theorem 14.

It is impossible that a field K have a separating transcendence basis over any subfield smaller than its maximal perfect subfield, as one sees by the following result.

THEOREM 19. *If a field K has a separating transcendence basis over a subfield L, then L contains the maximal perfect subfield of K.*

Proof. Let K^{p^∞} denote the maximal perfect subfield of K. If the conclusion were false, there would be a b in K^{p^∞} but not in L. All roots $b^{p^{-e}}$ lie in the perfect field $K^{p^\infty} \subset K$, so they all are separable over the extension $L(T)$ obtained from the given separating transcendence basis T for K/L. But $y = b$ is the root of some equation $g(y, T) = 0$ irreducible in the polynomial ring $L[y, T]$. Let t be a variable of T whose exponent p^e in g is as small as possible, so that no other t_i of T has exponent less than p^e, and let $T_0 = T - \{t\}$ be the set T with t deleted. Then $g(y, T) = g(y, t, T_0)$ is an irreducible separable equation for $y = b$ over $L(t^{p^e}, T_0^{p^e})$, with exponent 1 in the variable t^{p^e}. It follows (lemma below) that b is not separable over the smaller field $L(t^{p^{e+1}}, T_0^{p^e})$.

There is therefore an integer m and a corresponding set $S = T^{p^m}$ of transcendents such that all the roots $b^{p^{-n}}$ are separable over $L(S)$, while not all the roots $b^{p^{-n}}$ are separable over $L(S^p)$. Let $c = b^{p^{-d}}$ be one such inseparable root, so that c itself is not separable over $L(S^p)$, although $c^{p^{-1}}$ is separable over $L(S)$. If we apply the isomorphism $a \leftrightarrow a^p$, it follows that c is separable over $L^p(S^p)$, hence over the larger field $L(S^p)$, in contradiction to the choice of c. This contradiction establishes the theorem.

The lemma as to the inseparability of b used above is

LEMMA 5. *If a subset T of a field K is algebraically independent over a subfield L, and if an element b of K is a separable root $y = b$ of a polynomial $g(y, T)$ irreducible in the polynomial ring $L[y, T]$, then a variable t of T can appear in g with exponent 1 if and only if b is inseparable over $L(t^p, T - \{t\})$.*

The simple proof given in Mac Lane [5], Lemma II, §2, for the case when L is perfect, applies equally well for any field L. Easy consequences of the theorem are the following:

Corollary 1. *If a field K is obtained from a field L by successive transcendental and separable algebraic extensions, then the maximal perfect subfield of K is the maximal perfect subfield of L.*

Corollary 2. *If the elements of a set T are algebraically independent over a field L, then the intersection of all the fields $L(T^{p^e})$, for $e = 1, 2, \cdots$, is exactly the field L.*

9. **Pure forms.** The notion of a p-basis of a field can be used to show that Albert's results on pure null forms over a function field are in essence valid over an arbitrary coefficient field. We remark first that the *degree of imperfection* of a field K has been defined (Teichmüller [10]) to be the number of elements in a p-basis of K. A *pure form* f of degree q over K,

$$f(x_1, \cdots, x_m) = b_1 x_1^q + \cdots + b_m x_m^q \qquad (b_i \text{ in } K),$$

is a null form over K if $f(x_{10}, \cdots, x_{m0}) = 0$ for values $x_{10}, \cdots, x_{m0}$ not all zero in K.

Theorem 20. *Every pure form $f(x_1, \cdots, x_m)$ of degree $q = p^e$ over a field K of characteristic p is a null form if the number of variables m exceeds q^r, where r is the degree of imperfection of K. However, there exist non-null forms over K for every $m \leqq q^r$.*

The proof is exactly similar to that given by Albert ([2], Theorem 6).

We also prove here a property of p-bases which we have used elsewhere without proof ([6], Lemma 3).

Theorem 21. *If X and Y are disjoint subsets of a field K such that $X \cup Y$ is p-independent in K, and if $K(X^{p^{-\infty}})$ is obtained from K by adjoining all roots $x^{p^{-e}}$, for x in X and e an integer, then the set Y is p-independent in $K(X^{p^{-\infty}})$. Furthermore, if $X \cup Y$ is a p-basis of K, then Y is a p-basis of $K(X^{p^{-\infty}})$.*

Proof. If Y is not p-independent in $K(X^{p^{-\infty}})$, there is an element y in Y p-dependent on a finite subset Y_0 of Y not containing y. Hence for some integer e and some finite subset $X_0 \subset X$, $y \in K^p(X_0^{p^{-e}}, Y_0)$, so $y^{p^e} \in K^{p^{e+1}}(X_0, Y_0)$. This will lead to a contradiction on the degree of the field $L = K^{p^{e+1}}(X_0, Y_0, y)$ over $K^{p^{e+1}}$. For, on the one hand, the p-independence of $X \cup Y$ in K means that X_0, Y_0, and y together have degree $p^{\xi+\eta+1}$ over K^p, where ξ and η respectively denote the number of elements in X_0 and Y_0. Hence by an induction one finds

$$[L:K^{p^{e+1}}] = [K^{p^{e+1}}(X_0, Y_0, y):K^{p^{e+1}}] = p^{(\xi+\eta+1)(e+1)}.$$

On the other hand, $y^{p^e} \in K^{p^{e+1}}(X_0, Y_0)$, so that if we adjoin first all the elements of X_0 and Y_0, then finally the element y, we get for the same extension a degree

not more than $p^{(\xi+\eta)(e+1)+e}$, a contradiction. This establishes the p-independence of Y in $K(X^{p^{-\infty}})$; that it becomes a p-basis is readily verified from the definition.

10. **Fields without separating transcendence bases.** We now give a counter-example to the possible extension of Theorem 4 of the introduction; in that we show:

(i) There is a field M which does not have a separating transcendence basis over its maximal perfect subfield P, but which does have a finite transcendence degree t over that field P. Here t may be any specified integer $t \geqq 2$.

This example will also show:

(ii) The number of elements in a p-basis of a field K (the so-called degree of imperfection of K) is not always equal to the transcendence degree of K over its maximal perfect subfield.

Teichmüller, in [10], has also proved (ii) by the example in which K is a field of formal power series, where the transcendence degree in question is infinite. Our example yields a field K in which both the transcendence degree and the degree of imperfection are finite.

Let P be a perfect field and $Z = \{z_1, z_2, \cdots\}$, a denumerable set of quantities algebraically independent over P. Denote by $P(Z^{p^{-\infty}})$ the perfect field

$$P(Z^{p^{-\infty}}) = P(Z, Z^{p^{-1}}, Z^{p^{-2}}, \cdots). \tag{1}$$

Let y and u_0 be algebraically independent over $P(Z^{p^{-\infty}})$, and define quantities u_n recursively by

$$u_n = y^{p^{n-1}} + z_n u_{n-1} \qquad (n = 1, 2, \cdots). \tag{2}$$

The field which we use as an example is then

$$M = P(Z^{p^{-\infty}}, y, u_0, u_1^{1/p}, u_2^{1/p^2}, \cdots, u_n^{1/p^n}, \cdots).$$

By (2), $u_{n-1}^{p^{1-n}}$ can be expressed in terms of $u_n^{p^{1-n}}$, so M is the union of a tower of fields $M_0 \subset M_1 \subset M_2 \subset \cdots$, where

$$M_n = P(Z^{p^{-\infty}}, y, u_n^{1/p^n}) \qquad (n = 0, 1, 2, \cdots).$$

From equations (2) we observe that

$$P(z_1, \cdots, z_n, y, u_0) = P(z_1, \cdots, z_n, y, u_n) \tag{3}$$

and hence that $z_1, \cdots, z_n, y, u_n$ are algebraically independent over P. Furthermore

$$M_0 = P(Z^{p^{-\infty}}, y, u_n), \qquad M_n = M_0(u_n^{p^{-n}}), \tag{4}$$

so that M_n is a purely inseparable extension of M_0 of degree p^n. Therefore the necessary condition of Theorem 6 applied to the transcendence basis $T = \{y, u_0\}$ for M proves the following result:

LEMMA 6. *The field M does not have a separating transcendence basis over $P(Z^{p^{-\infty}})$.*

The extended field $M(y^{1/p})$ would by the equations (2) also contain each $u_{n-1}^{p^{-n}}$, so that we can assert

LEMMA 7. *The field M has a p-basis consisting of one element y and a transcendence basis $\{y, u_0\}$ over the perfect subfield $P(Z^{p^{-\infty}})$.*

Thus our example has the properties (i) and (ii) stated above, provided we can prove that $P(Z^{p^{-\infty}})$ is the maximal perfect subfield of M. This we now do.

LEMMA 8. *An element b of M is in M_n if and only if b^{p^n} is separable over M_0.*

Proof. By (4), all elements b of M_n certainly have the property stated. Conversely, suppose b^{p^n} separable over M_0, and take $k \geqq n$ so large that $b \epsilon M_k = M_0(u_k^{p^{-k}})$. In such a field M_k it is known that any element b with b^{p^n} separable over the base field M_0 must be[22] in $M_0(u_k^{p^{-n}})$. If $k > n$, (2) yields $u_k^{p^{-n}} = y^{p^{k-n-1}} + z_k^{p^{-n}} u_{k-1}^{p^{-n}}$, hence $M_0(u_k^{p^{-n}}) = M_0(u_{k-1}^{p^{-n}})$. Therefore, by induction, $b \epsilon M_0(u_n^{p^{-n}}) = M_n$, as asserted.

Consider now any element a in the maximal perfect subfield M^{p^∞} of M. The expansion for a in terms of the generators of M can involve but a finite number of u's, but a finite number of z's, and but a finite number of p^n-th roots of z's. Hence we can find a power $c = a^{p^n}$, also in M^{p^∞}, and an integer m such that $c \epsilon P(Z_m, y, u_0)$, where Z_m is a finite subset

$$Z_m = \{z_1, \cdots, z_m\}. \tag{5}$$

For each e, $c = c_e^{p^e}$ for some element c_e of M, while by Lemma 8, $c_e \epsilon M_e$. Thus

$$c \epsilon P(Z_m, y, u_0), \qquad c \epsilon M_e^{p^e} \qquad (e = 0, 1, 2, \cdots). \tag{6}$$

This situation will now be simplified by showing that only a finite number of z's need be used in these fields M_e, provided $e \geqq m$. Note that $M_e^{p^e} = P(Z^{p^{-\infty}}, y^{p^e}, u_e)$.

For each $e \geqq m$ pick the smallest n such that the first n z's suffice to make $c \epsilon M_e^{p^e}$; i.e., such that $c \epsilon P(Z_n^{p^{-\infty}}, y^{p^e}, u_e)$. If $n > e$, c is not in the field $R = P(Z_{n-1}^{p^{-\infty}}, y^{p^e}, u_e)$, but is in $R(z_n^{p^{-t}})$ for some t. As in the remark following (3), z_n is transcendental over $P(Z_{n-1}, y, u_e)$, hence over R. Thus c in the purely transcendental extension $R(z_n^{p^{-t}})$ must itself be transcendental over R. On the other hand, (6) and (4) state that

$$c \epsilon P(Z_m, y, u_0) \subset P(Z_e, y, u_0) = P(Z_e, y, u_e) \subset P(Z_{n-1}, y, u_e),$$

and the last of these fields is algebraic over $R = P(Z_{n-1}^{p^{-\infty}}, y^{p^e}, u_e)$. Therefore c is algebraic over R, a contradiction from which we conclude that

$$c \epsilon P(Z_e^{p^{-\infty}}, y^{p^e}, u_e) \qquad (e = m, m+1, \cdots).$$

If the expression for c in terms of these generators actually involves some roots of Z_e, pick $s > 0$ so small that c is in $N^* = P(Z_e^{p^{-s}}, y^{p^e}, u_e)$ but not in $N =$

[22] Mac Lane [5], Lemma 6.1: *If u is transcendental over F, then the elements b of $F(u^{p^{-\infty}})$ with b^{p^e} separable over $F(u)$ all lie in $F(u^{p^{-e}})$.*

$P(Z_e^{p^{-s+1}}, y^{p^e}, u_e)$. The extension N^*/N is given by a tower $N \subset N_1 \subset N_2 \subset \cdots \subset N^*$, where

$$N_0 = N, \qquad N_i = N(z_1^{p^{-s}}, \cdots, z_i^{p^{-s}}) \qquad (i = 0, \cdots, e).$$

For some i, c is in N_i but not in N_{i-1}. Since $N^{*p} \subset N$, c is purely inseparable over N_{i-1}, and the two extensions $N_{i-1}(c)$ and $N_i = N_{i-1}(z_i^{p^{-s}})$, each of degree p, must be equal. Therefore $z_i^{p^{-s}} \in N_{i-1}(c)$. But now set $F = P([Z - \{z_i\}]^{p^{-\infty}}, y, u_0)$, so that $N_{i-1} \subset F(z_i^{p^{-s+1}})$, while by (6), $c \in P(Z_m, y, u_0) \subset F(z_i^{p^{-s+1}})$. Therefore $z_i^{p^{-s}} \in F(z_i^{p^{-s+1}})$. This conclusion is a contradiction because z_i is a quantity transcendental over F. Hence we have $s = 0$, and

$$c \in P(Z_e, y^{p^e}, u_e) \qquad (e = m, m + 1, \cdots). \tag{7}$$

Combining (6) with (7), we next aim to prove that c is in each of the fields

$$D_{em} = P(Z_m, y^{p^e}, u_m^{p^{e-m}}) \qquad (e = m, m + 1, \cdots). \tag{8}$$

Note that D_{ee} is the field $P(Z_e, y^{p^e}, u_e)$ of (7). A preliminary is

Lemma 9. *Each D_{em} with $e \geqq m$ is relatively algebraically closed* (see footnote 13) *in D_{ee}.*

Proof. First simplify the notation of (8) thus:

$$D_{en} = P(Z_n, y', v), \qquad y' = y^{p^e}, \qquad v = u_n^{p^{e-n}}, \tag{9}$$

$$D_{en+1} = P(Z_n, z, y', u), \qquad z = z_{n+1}, \qquad u = u_{n+1}^{p^{e-n-1}}. \tag{10}$$

In terms of these quantities u and v the defining equation (2) becomes

$$u^p = y' + z^{p^{e-n}} v. \tag{11}$$

Hence $D_{en} \subset D_{en+1}$, according to (9) and (10), and we can go from D_{en} to D_{ee} by a tower

$$D_{en} \subset D_{en+1} \subset D_{en+2} \subset \cdots \subset D_{ee}.$$

For the lemma it therefore suffices to prove D_{en} relatively algebraically closed in D_{en+1} for all $n < e$.

By (9) and (10) $D_{en+1} = D_{en}(z, u)$, where, as in (3), z is transcendental over D_{en}, while u is inseparable and algebraic over $D_{en}(z)$ in accord with (11). If D_{en} is not relatively algebraically closed in D_{en+1}, pick b in $D_{en+1} - D_{en}$ and algebraic over D_{en}. Then b^p is in $D_{en}(z)$ and is algebraic over D_{en}, so b^p is in D_{en}. Therefore b is purely inseparable over D_{en} and also over $D_{en}(z)$, and we must have $D_{en+1} = D_{en}(z, b)$. Therefore u is a rational function $g(z)/h(z)$ of z with coefficients in $D = D_{en}(b)$. We can assume that $g(0)$ and $h(0)$ are not both 0. This value of u in the defining equation (11) for u gives an identity

$$[h(z)]^p[y' + z^{p^{e-n}} v] = [g(z)]^p \tag{12}$$

in z over D. By setting $z = 0$, we find $y' = [g(0)/h(0)]^p$, with neither $h(0)$ nor $g(0)$ zero. This means that y' is in D^p, hence that $(y')^{1/p}$ is in D. A similar

argument on the terms of highest degree in (12) proves that $v^{1/p} \in D$. However, in D_{en} of (9), Z_n, y' and v are algebraically independent over P, so that y', v are p-independent in D_{en}. This means that the extension $D_{en}((y')^{1/p}, v^{1/p})$ has degree p^2 over D_{en}, although we have just shown this extension to be contained in D, of degree p over D_{en}. This contradiction establishes the desired relative algebraic closure.

Lemma 10. *For* $e \geqq m$, $c \in D_{em}$.

Proof. By (6), $c \in P(Z_m, y, u_0) = P(Z_m, y, u_m)$, a field which is certainly algebraic over D_{em} of (8). On the other hand $c \in D_{ee} = P(Z_e, y^{p^e}, u_e)$, by (7), and this field by the previous lemma contains no elements algebraic over D_{em} except the elements of D_{em} themselves. Hence we get the conclusion.[23]

If we put $L = P(Z_m)$, $\mu = e - m$, Lemma 10 states that c is in each of the fields $D_{em} = L((y^{p^m})^{p^\mu}, u_m^{p^\mu})$. Here y^{p^m} and u_m are algebraically independent over L, so the intersection of all these fields, for $\mu = 1, 2, \cdots$, is known by Corollary 2 to Theorem 19 of §8 to be L itself. Therefore $c \in P(Z_m)$, and the original element $a = c^{p^{-r}}$ of the maximal perfect subfield is therefore in $P(Z_m^{p^{-\infty}})$. We have thus completed our example by proving

Lemma 11. *The field M has the maximal perfect subfield $P(Z^{p^{-\infty}})$.*

This field M thus has a transcendence degree 2 over its maximal perfect subfield. A field with analogous properties but with any desired transcendence degree $t \geqq 2$ over its maximal perfect subfield, is, by Corollary 1 of Theorem 19, the field $M^* = M(T)$, where T is a set of $t - 2$ elements algebraically independent over M.

Bibliography

1. A. A. Albert, *Modern Higher Algebra*, Chicago, 1937.
2. A. A. Albert, *Quadratic null forms over a function field*, Annals of Mathematics, vol. 39(1938), pp. 494–505.
3. M. Becker and S. Mac Lane, *The minimum number of generators for inseparable algebraic extensions*. To be published.
4. H. Hasse and F. K. Schmidt, *Die Struktur diskret bewerteter Körper*, Journal für die Mathematik, vol. 170(1934), pp. 4–63.
5. S. Mac Lane, *Steinitz field towers for modular fields*. Forthcoming in the Transactions of the American Mathematical Society.
6. S. Mac Lane, *Subfields and automorphism groups of p-adic fields*, Annals of Mathematics, vol. 40(1939), pp. 423–442.
7. S. Mac Lane, *A lattice formulation for transcendence degrees and p-bases*, this Journal, vol. 4(1938), pp. 455–468.
8. F. K. Schmidt, *Allgemeine Körper im Gebiet der höheren Kongruenzen*, Dissertation, Erlangen, 1925.
9. E. Steinitz, *Algebraische Theorie der Körper*, edited by R. Baer and H. Hasse, Berlin, 1930.

[23] The arguments of Lemmas 8 and 9 are the crux of this example. As given, they depend essentially upon the algebraic independence of the z's. This is the inner reason for the complicated structure of the maximal perfect field $P(Z^{p^{-\infty}})$ used for this example.

10. O. TEICHMÜLLER, *p-Algebren*, Deutsche Mathematik, vol. 1(1936), pp. 362–388.
11. O. TEICHMÜLLER, *Diskret bewertete perfekte Körper mit unvollkommenen Restklassenkörper*, Journal für die Mathematik, vol. 176(1937), pp. 126–140.
12. B. L. VAN DER WAERDEN, *Moderne Algebra*, vol. I, first edition, Berlin, 1930.
13. B. L. VAN DER WAERDEN, *Zur algebraischen Geometrie* XIV, *Schnittpunktszahlen von algebraischen Mannigfaltigkeiten*, Mathematische Annalen, vol. 115(1938), pp. 619–644.

HARVARD UNIVERSITY.

OSCILLATING FUNCTIONS

By R. P. Boas, Jr.

1. **Introduction.** We may say that a function $f(x)$ is monotonic at the point x_0 if there is a positive δ such that, whenever $x_0 - \delta < x_1 \leqq x_0 \leqq x_2 < x_0 + \delta$, either $f(x_1) \leqq f(x_0) \leqq f(x_2)$ or $f(x_1) \geqq f(x_0) \geqq f(x_2)$; a function monotonic at x_0 is not necessarily monotonic in any interval containing x_0. There are then several senses in which a continuous function $f(x)$ may be said to be everywhere oscillating: $f(x)$ may be monotonic in no interval, almost nowhere monotonic (i.e., monotonic at most at the points of a set of measure zero), monotonic at most at the points of a countable set, or monotonic at no point. The most natural questions of the existence of functions monotonic in no interval, belonging to more or less restricted classes, are settled by the functions constructed by P. Köpcke and A. Denjoy,[1] which are monotonic in no interval, and not only absolutely continuous, but differentiable at every point, with bounded derivatives. P. Hartman and R. Kershner[2] have recently given a simple construction of an absolutely continuous function which is monotonic in no interval.

It is evident that an absolutely continuous function cannot be almost nowhere monotonic, since it is surely monotonic at the points of the set where its derivative is not zero. Similarly, it is clear that a function which almost everywhere fails to have a finite derivative is almost nowhere monotonic, since by a well known theorem,[3] such a function will almost everywhere have one of its upper Dini derivatives $+\infty$, and one of its lower Dini derivatives $-\infty$. These considerations tell us nothing about the existence of a continuous function of bounded variation, almost nowhere monotonic; in this note such a function will be constructed. A continuous function $f(x)$ of bounded variation must, however, be monotonic at the points of an uncountable set.[4] For, let the curve $y = f(x)$ $(0 \leqq x \leqq 1)$ have the parametric representation $x = x(s)$, $y = y(s)$ $(0 \leqq s \leqq l, l > 1)$, where s is the arc length. Then[5] $x'(s)^2 + y'(s)^2 = 1$

Received January 16, 1939. The author is a National Research Fellow.

[1] A. Denjoy, *Sur les fonctions dérivées sommables*, Bulletin de la Société Mathématique de France, vol. 43(1915); pp. 161–248; pp. 210 ff. Denjoy gives a critique of Köpcke's construction (pp. 228 ff.).

[2] P. Hartman and R. Kershner, *The structure of monotone functions*, American Journal of Mathematics, vol. 59(1937), pp. 809–822; p. 817.

[3] E. W. Hobson, *The Theory of Functions of a Real Variable and the Theory of Fourier's Series*, vol. 1, 1927, p. 400.

[4] I am indebted to A. P. Morse for this remark.

[5] See, e.g., F. Riesz, *Sur l'existence de la dérivée des fonctions monotones et sur quelques problèmes qui s'y rattachent*, Acta Litterarum ac Scientiarum Regiae Universitatis Hungaricae Francisco-Josephinae, Sectio Scientiarum Mathematicarum [Szeged], vol. 5(1930–32), pp. 208–221; p. 216.

for almost all s. Moreover, $y'(s) \neq 0$ for a set of values s of positive measure, since otherwise we should have $x'(s) = 1$ for almost all s, and therefore $l = \int_0^l x'(s)\,ds \leqq x(l) - x(0) = 1$. Hence, for a set of values of s of positive measure, and consequently for an uncountable set of values of x, $f'(x) = y'(s)/x'(s) \neq 0$; when $f'(x_0) \neq 0$, $f(x)$ is evidently monotonic at x_0.

It is reasonable to suppose that "most" functions are everywhere oscillating. We shall make this statement precise in terms of the category of various classes of everywhere oscillating functions in various spaces. Again, several statements are trivial consequences of known results. Thus from a theorem of J. C. Oxtoby[6] it follows immediately that functions monotonic in no interval form a residual G_δ set[7] in the space AC of absolutely continuous functions. Since nowhere-differentiable functions form a residual set in the spaces H^α of functions satisfying Lipschitz conditions of order α $(0 < \alpha < 1)$,[8] the nowhere monotonic functions of these spaces also form residual sets.

We shall show that in the space CBV of functions of bounded variation the functions monotonic in no interval form a residual G_δ set; the method could easily be adapted to establish the same result for the spaces AC and H^α as well. A simpler, but closely related, result is that the set of functions of CBV, monotonic at a specified point, is of the first category.

Let R be the space of integrals of essentially bounded functions on $(0, 1)$, vanishing at the origin, with the natural norm[9] $\| f(x) \| = \sup^\circ | f'(x) |$. It is clear that the set O of elements of R, monotonic in no interval, is not residual;[10] neither is its complement (once we know that O is not empty). We can, however, by using a different metric, establish by a simple category argument the existence of a function, monotonic in no interval, with a bounded derivative; regarded as a construction, this of course establishes less than the constructions of Köpcke and Denjoy.

2. **The oscillating functions of** CBV.

The elements of CBV are continuous functions $x(t)$ of bounded variation on $(0, 1)$, normalized by the condition $x(0) = 0$, and with norm $\| x \| = \int_0^1 | dx(t) |$. With the obvious definitions

[6] J. C. Oxtoby, *The category and Borel class of certain subsets of* $\mathfrak{L}_p$, Bulletin of the American Mathematical Society, vol. 43(1937), pp. 245–248; Theorem 5.

[7] A set of the first category is the sum of a countable number of nowhere dense sets. A residual set is the complement of a set of the first category. A G_δ set is the intersection of a countable number of open sets; an F_σ is the complement of a G_δ.

[8] H. Auerbach and S. Banach, *Über die Höldersche Bedingung*, Studia Mathematica, vol. 3(1931), pp. 180–188.

[9] The superscript ° attached to a symbol indicates the disregard of certain exceptional sets of measure zero. Thus sup° means "essential least upper bound" (= "true max", "ess. sup.", etc.); =° means "equals almost everywhere"; etc. (Notation suggested by F. Smithies.)

[10] It is not even everywhere dense.

of the operations in the space, CBV is a Banach space. Let O be the set of elements of CBV, monotonic in no subinterval of $(0, 1)$.

THEOREM 1. *The set O is a residual G_δ set in CBV.*

Let $C(O)$ be the complement of O. Let $\{I_n\}$ be the sequence of subintervals of $(0, 1)$ with rational endpoints, and denote by E_n the set of elements of CBV which are monotonic in I_n.[11] Evidently each E_n is closed (since if $x_k \to x$ in CBV, $x_k(t) \to x(t)$ uniformly in $0 \leqq t \leqq 1$). Moreover, $C(O) = \sum_{n=1}^{\infty} E_n$. To establish Theorem 1, then, we have to show that each E_n is nowhere dense.

Since each E_n is closed, it is enough to show that for any $x \in CBV$, and η $(0 < \eta < 1)$, we can construct $y \in C(E_n)$ such that $\| y - x \| < \eta$; for, this shows that $C(E_n)$ is everywhere dense, and a closed set with a dense complement is nowhere dense.

We introduce an auxiliary function $z(t; \tau, \gamma, \beta)$ defined for $0 \leqq t \leqq 1$, $0 \leqq \tau < \tau + \gamma < 1$, $\beta > 0$ to be continuous in $(0, 1)$; zero in $(0, \tau)$ and $(\tau + \gamma, 1)$; β at $t = \tau + \frac{1}{2}\gamma$; and linear in $(\tau, \tau + \frac{1}{2}\gamma)$ and in $(\tau + \frac{1}{2}\gamma, \tau + \gamma)$. Then $z(t; \tau, \gamma, \beta) \in CBV$ and $\| z \| = 2\beta$.

If $x(t)$ is not monotonic in I_n, we take $y(t) \equiv x(t)$.

If $x(t)$ is constant in I_n, set $y(t) = x(t) + z(t; \lambda, \delta, \delta)$, where λ is the midpoint of I_n, $\delta < \frac{1}{2}\eta$, and 2δ is less than the length of I_n. Then $\| y - x \| = 2\delta < \eta$, and $y(t)$ is not monotonic in I_n.

If $x(t)$ is monotonic and not constant in $I_n = (t', t'')$, there is a point $s \geqq t'$ such that $x(t) = x(t')$ in (t', s), but $x(t) \neq x(t')$ for $s < t \leqq t''$. We take a number $A > 0$ and a point τ such that $s < \tau < t''$ and $| x'(\tau) | < A$. We set $y(t) = x(t) \mp z(t; \tau, 2\delta, 3A\delta)$, where $6A\delta < \eta$, $\tau + 2\delta \leqq 1$, and the upper or lower sign is taken according as $x(t)$ is increasing or decreasing[12] in I_n.

Then $\| y - x \| = 6A\delta < \eta$. Suppose, for definiteness, that $x(t)$ increases in I_n. Then $x'(\tau) < A$, and hence, for sufficiently small positive ζ, $x(\tau + \zeta) - x(\tau) < 2A\zeta$. For $0 < \zeta < \delta$, $z(\tau + \zeta; \tau, 2\delta, 3A\delta) - z(\tau; \tau, 2\delta, 3A\delta) = 3A\zeta$. Therefore $y(\tau + \zeta) - y(\tau) < 2A\zeta - 3A\zeta < 0$. On the other hand, $y(t) \equiv x(t)$ in (s, τ), and $x(t)$ increases there and is not constant. Hence $y(t)$ is not monotonic in I_n. Similar reasoning applies if $x(t)$ decreases in I_n; this completes the proof of Theorem 1.

3. **Functions monotonic at a particular point.** Let us say that $x(t)$ increases on the right at t_0 if there is a positive δ such that $x(t_0) \leqq x(t)$ when $t_0 \leqq t \leqq t_0 + \delta$. We have the following theorem.

THEOREM 2. *For any t_0, $0 < t_0 < 1$, the sets $I(t_0)$, $D(t_0)$ of elements of CBV which, respectively, increase or decrease on the right at t_0 are F_σ sets of first category.*

[11] This choice of sets E_n was suggested by the referee, and leads to a considerable simplification of my original proof.

[12] By "increasing" we mean, throughout, "increasing in the weak sense" ("non-decreasing"). Similarly for "decreasing".

Theorem 2 is also true if CBV is replaced by AC or H^α $(0 < \alpha < 1)$; the proof requires only slight modifications.

Theorem 2 leads to a simplified proof of the category result of Theorem 1. Let $\{t_n\}$ $(n = 1, 2, \cdots)$ be a countable set of points dense in $(0, 1)$. Let $H_1, H_2, \cdots$ denote the intervals of the countable set $(k/m, (k + 1)/m)$ $(m = 1, 2, \cdots; k = 0, 1, \cdots, m - 1)$. If $x \in C(O)$, x is monotonic in some H_k, and consequently increasing or decreasing on the right at every point t_n in this H_k. We therefore have

$$C(O) \subset \sum_{k=1}^{\infty} \prod_{t_n \in H_k} \{I(t_n) + D(t_n)\},$$

so that $C(O)$ is a set of first category.

We now prove Theorem 2 for $I(t_0)$. We have $I(t_0) = \sum_{n=1}^{\infty} E_n$, where E_n is the set of x such that $x(t) - x(t_0) \geqq 0$ for $t_0 \leqq t < t_0 + n^{-1}$. Evidently each E_n is closed, and Theorem 2 follows if we show that the complement of each E_n is everywhere dense; for then each E_n is nowhere dense.

Let x be any element of CBV, and let $\eta > 0$ be arbitrary. We construct, for each n, $y \in C(E_n)$ with $\| y - x \| \leqq \eta$. To do this, choose $\epsilon > 0$ so that $2\epsilon < n^{-1}$ and $x(t) \leqq x(t_0) + \frac{1}{3}\eta$ for $t_0 \leqq t \leqq t_0 + 2\epsilon$. Then if $y(t) = x(t) - z(t; t_0, 2\epsilon, \frac{1}{2}\eta)$, $\| y - x \| = \eta$; and $y(t_0 + \epsilon) = x(t_0 + \epsilon) - \frac{1}{2}\eta < x(t_0) = y(t_0)$, so that $y \in C(E_n)$.

4. **Oscillating functions with bounded derivatives.** Let S be the space of measurable functions $x(t)$ $(0 \leqq t \leqq 1)$ such that $\sup^\circ | x(t) | \leqq 1$; S is a complete metric space under the "L metric" $(x, y) = \int_0^1 | x(t) - y(t) | \, dt$. Now let T be the complete metric space whose elements are absolutely continuous functions $x(t)$ with $x(0) = 0$ and $x'(t) \in S$, and with the distance between x and y defined as (x', y'). (Note that the completeness of T is ensured by the *uniform* essential boundedness of the $x'(t)$.) Then there is an obvious isometric relation between T and S in which the set O of functions of T, monotonic in no interval, corresponds to the set of functions of S which are not of essentially fixed sign in any interval.

We shall prove

THEOREM 3. *The set O is a residual G_δ set in T (and consequently not empty[13]).*

Let $\{I_n\}$ be the sequence of "rational intervals", as in the proof of Theorem 1, and let E_n be the set of $x \in S$ such that, in I_n, $x(t) \geqq^\circ 0$ or $x(t) \leqq^\circ 0$. Then each E_n is closed.

Consider any $x \in E_n$, and any $\eta > 0$. If $x(t) =^\circ 0$ in I_n, set $y(t) = x(t)$ if

[13] By Baire's theorem: A complete metric space is not of the first category. (See, e.g., C. Kuratowski, *Topologie* I, 1933, pp. 204–205.)

t is not in I_n; and $y(t) = \eta/\lambda$ in half of I_n, $y(t) = -\eta/\lambda$ in the other half, where λ is the length of I_n; then $y \in C(E_n)$ and $\| y - x \| = \eta$.

If $x(t) >^\circ 0$ or $x(t) <^\circ 0$ in I_n, take a set $J \subset I_n$, of measure ϵ $(0 < \epsilon < \frac{1}{2}\eta)$, such that $| x(t) | >^\circ 0$ in J and in $I_n - J$; and define $y(t) = x(t)$ if $t \in (0, 1) - J$, $y(t) = -x(t)$ if $t \in J$. Then $y \in C(E_n)$; and $\| y - x \| \leqq 2\epsilon < \eta$.

Thus E_n is closed and contained in the closure of its complement, and is consequently nowhere dense. Hence $\sum_{n=1}^{\infty} E_n$ is an F_σ set of first category, and Theorem 3 follows.

5. **An oscillating function of** *CBV*. Our aim in this section is to construct an almost nowhere monotonic function of *CBV*; the function which we shall construct will actually have the property that, except at the points of a set of measure zero, it is monotonic neither on the right nor on the left.

We denote by $| E |$ the measure of the set E. Let $h_\alpha(x)$ be the function which increases on $(0, 1)$ from 0 to $\alpha > 0$, and which is constant on each complementary interval of the Cantor ternary set (of measure zero). Let $g_\alpha(x)$ denote the function defined as

$$\begin{array}{rl} h_\alpha(x), & 0 \leqq x \leqq 1, \\ h_\alpha(2 - x), & 1 \leqq x \leqq 2, \\ -h_\alpha(x - 2), & 2 \leqq x \leqq 3, \\ -h_\alpha(4 - x), & 3 \leqq x \leqq 4, \\ h_\alpha(x - 4), & 4 \leqq x \leqq 5, \\ h_\alpha(6 - x), & 5 \leqq x \leqq 6. \end{array}$$

Thus $g_\alpha(x)$ increases from 0 to α on $(0, 1)$, decreases to $-\alpha$ on $(1, 3)$, increases to α on $(3, 5)$, and decreases to 0 on $(5, 6)$. We shall construct our function $f(x)$ (on $(0, 6)$ instead of on $(0, 1)$) by iterating a process of inserting suitably reduced copies of functions $g_\alpha(x)$ in the intervals in which $g_1(x)$ is constant.

Let $\epsilon_n > 0$, $\epsilon_1 = 1$, $\sum_{n=1}^{\infty} \epsilon_n < \infty$. We proceed to define a sequence of functions $f_n(x)$.

We set $f_0(x) \equiv 0$, and $f_1(x) \equiv g_1(x)$. Then $f_1(x)$ is constant in each (open) interval of a countable set of intervals $I(1, k)$ $(k = 1, 2, \cdots)$; the closed set E_1 complementary to the sum of the $I(1, k)$ has measure zero. We now define $f_n(x)$ by induction.

Suppose that $f_m(x)$ has been constructed for $m \leqq n - 1$ $(n \geqq 2)$; that $f_m(x)$ is constant in each interval of a set $\{I(m, k)\}$, where the closed set E_m complementary to $\sum_{k=1}^{\infty} I(m, k)$ has measure zero; and that each $I(m, k)$ is contained in some $I(m - 1, j)$. Suppose further that $f_m(x) > f_{m-1}(x)$ in the first and third

thirds of each $I(m - 1, k)$, while $f_m(x) < f_{m-1}(x)$ in the middle third of each $I(m - 1, k)$.

We set $f_n(x) = f_{n-1}(x)$ for $x \in E_{n-1}$. Let $I(n - 1, k) = (a_k, b_k)$, and let $f_{n-1}(x) = \lambda_k$ in $I(n - 1, k)$. We enumerate the intervals $I(n - 1, k)$ by the rules that of two intervals of unequal length, the longer precedes; and of two intervals of equal length, the left-hand one precedes. Suppose that

$$|I(n - 1, r - 1)| > |I(n - 1, r)| = |I(n - 1, r + 1)| = \cdots = |I(n - 1, s)| = \delta > |I(n - 1, s + 1)|.$$

For $r \leqq k \leqq s$, we set

$$f_n(x) = g_\alpha\left(6\frac{x - a_k}{\delta}\right) + \lambda_k, \qquad x \in I(n - 1, k),$$

where $\alpha < \eta_k$ $(k = r, r + 1, \cdots, s)$; $\sum_{k=1}^{\infty} \eta_k < \epsilon_n$; and α is so small that if $I(n - 1, k)$ is in the first or third third of the $I(n - 2, j)$ which contains it, the minimum in $I(n - 1, k)$ of $f_n(x)$ is greater than the (constant) value of $f_{n-2}(x)$ in $I(n - 2, j)$, while if $I(n - 1, k)$ is in the middle third of $I(n - 2, j)$, the maximum in $I(n - 1, k)$ of $f_n(x)$ is less than the value of $f_{n-2}(x)$ in $I(n - 2, j)$.

Then $f_n(x)$ is constant in each interval $I(n, k)$ of a set of total measure one; $f_n(x) > f_{n-1}(x)$ in the first and third thirds of each $I(n - 1, k)$, while $f_n(x) < f_{n-1}(x)$ in the middle third; and each $I(n, k)$ is inside some $I(n - 1, j)$.

We now observe some properties of the $f_n(x)$, easily established by induction. Each $f_n(x)$ is continuous. The total variation on $(0, 6)$ of $f_n(x) - f_{n-1}(x)$ is not greater than $6\epsilon_n$. In two intervals $I(n - 1, k)$ of equal length, the graphs of $f_n(x) - f_{n-1}(x)$ are congruent. Moreover, in an $I(n - 1, k)$ the graph of $f_{n-1}(x)$ is a horizontal straight line, L; the graphs of $f_n(x)$ and $f_{n+1}(x)$ are above L in the first and third thirds of $I(n - 1, k)$, and below L in the middle third.

Now the sequence $\{f_n(x)\}$ is evidently uniformly convergent; and since the total variation of $f_n(x)$ is less than $6\sum_{n=1}^{\infty} \epsilon_n$, the limit, $f(x)$, is an element of CBV.

Let $E = \sum_{n=1}^{\infty} E_n$; then $|E| = 0$. If $t \in C(E)$, there is a unique set of intervals $I(1, k_1) \supset I(2, k_2) \supset \cdots$, such that $I(n, k_n) \to t$. Except for a set F of measure zero, the points $t \in C(E)$ have the property that if $I(1, k_1) \supset I(2, k_2) \supset \cdots \to t$, the sequences of indices k_n for which t is in the first, second, or third third of $I(n, k_n)$ are all infinite. In fact, $F = \sum_{m=1}^{\infty} (F_m^1 + F_m^2 + F_m^3)$, where F_m^σ is the set of points t such that for $n \geqq m$, t is not in the σ-th third of $I(n, k_n)$. Consider, say, an F_m^1. It is contained in the set of all right-hand two-thirds of intervals $I(m, k)$, i.e., in a set of measure $\frac{2}{3}\cdot 6$. Also, every point of F_m^1 is in a right-hand two-thirds of an interval $I(m + 1, j)$ which is in turn contained in a right-hand two-thirds of an $I(m, k)$; hence F_m^1 is contained in a set of measure $\frac{2}{3}\cdot\frac{2}{3}\cdot 6$. Continuing in this way, we see that $|F_m^1| = 0$; and similarly that $|F_m^2| = |F_m^3| = 0$.

Let t be an arbitrary point of $G = C(E) - F$; then $| G | = 1$. We shall show that $f(x)$ is not monotonic on the right at t; a similar proof would show that $f(x)$ is not monotonic on the left at t.

We have $t = \lim_{n\to\infty} I(n, k_n)$. Let $I(n, k_n) = J_n + K_n + L_n$ be a decomposition of $I(n, k_n)$ into successive thirds. Let $\epsilon > 0$ be arbitrary, and let n_0 be so large that $2 | I(n, k_n) | < \epsilon$ for $n \geqq n_0$. For an infinite number of values $n \geqq n_0$, $I(n + 1, k_{n+1}) \subset J_n$; choose such an n, and fix it. The function $f_{n+1}(x)$ is constant in various subintervals of K_n, and at least one of these subintervals is of the same length as $I(n + 1, k_{n+1})$; choose from these subintervals one in which the maximum of $f_{n+1}(x)$ is smallest, and call it H_{n+1}. Our definition of $f_{n+1}(x)$ was such that

$$\inf_{x \epsilon J_n} f_{n+1}(x) - \sup_{x \epsilon H_{n+1}} f_{n+1}(x) = \theta > 0.$$

Now consider $I(n + 2, k_{n+2}) \subset I(n + 1, k_{n+1}) \subset J_n$. H_{n+1} contains some interval $I(n + 2, j)$ of length $| I(n + 2, k_{n+2}) |$; call this interval H_{n+2}. Since $f_{n+2}(x)$ is obtained from $f_{n+1}(x)$ by adding congruent functions in all intervals $I(n + 2, j)$ of the same length, we have

$$\inf_{x \epsilon I(n+2, k_{n+2})} f_{n+2}(x) - \sup_{x \epsilon H_{n+2}} f_{n+2}(x) \geqq \theta.$$

We continue in the same way, choosing H_{n+p} $(p = 1, 2, \cdots)$ as a subinterval of H_{n+p-1}, so that $| H_{n+p} | = | I(n + p, k_{n+p}) |$. The H_{n+p} converge to a point t', and $0 < t' - t < \epsilon$; we have $f(t) - f(t') \geqq \theta > 0$. For any $\epsilon > 0$, we can find such a t' and θ, and hence $f(x)$ cannot increase on the right at t.

We then carry out the same process, starting with a value of n such that $I(n + 1, k_{n+1}) \subset K_n$, defining t' as a limit of a sequence of intervals $I(n + p, k_{n+p}) \subset L_n$. In this case we obtain $f(t) < f(t')$, where $0 < t' - t \leqq \epsilon$, and consequently $f(x)$ cannot decrease on the right at t.

CAMBRIDGE, ENGLAND.

A DIFFERENTIAL EQUATION FOR ORTHOGONAL POLYNOMIALS

By J. Shohat

Introduction. Any sequence of orthogonal polynomials[1] (OP) $\{\Phi_n(x)\}$ satisfies, as is known, a linear homogeneous difference equation of second order

$$(1)\qquad \Phi_n(x) - (x - c_n)\Phi_{n-1}(x) + \lambda_n\Phi_{n-2}(x) = 0 \qquad (n \geqq 2; \Phi_0 = 1, \Phi_1 = x - c_1),$$

where λ_n, c_n are constants, $\lambda_n > 0$. On the other hand, the classical OP of Hermite, Laguerre and Jacobi (special cases: Legendre's and trigonometrical polynomials) satisfy, in addition, a homogeneous linear differential equation of the following type (M, p. 33):

$$(2)\qquad A\Phi_n''(x) + B\Phi_n'(x) + C_n\Phi_n(x) = 0,$$

where A, B are polynomials in x, independent of n, of degrees not exceeding 2 and 1, respectively, and C_n is a constant depending on n.

The importance of differential equations in the study of OP needs no further emphasis. Thus, it is natural to seek to find other classes of OP for which a differential equation of this type exists, namely:

$$(3)\qquad A_n\Phi_n''(x) + B_n\Phi_n'(x) + C_n\Phi_n(x) = 0,$$

where A_n, B_n, C_n are polynomials in x, *each of fixed degree independent of* n, with coefficients eventually depending on n. In a note in the Comptes Rendus[2] the author has shown the existence of (3) for a certain general class of OP. The method employed, following Laguerre, yields rather an "existence proof" and is not readily applicable to the actual construction of the polynomials A_n, B_n, C_n.

The object of the present paper is to develop a new simple method for the effective construction of the differential equation (3) for an extended class of OP, of the same general type as in the note just cited. In application to the classical and other OP, this method yields, as by-products, many of their properties, old and new.

The method used is of a very elementary character.

Received January 23, 1939; presented to the American Mathematical Society, December 30, 1938.

[1] $\Phi_n(x) = x^n - S_nx^{n-1} + d_{n,n-2}x^{n-2} + \cdots$; $\{\varphi_n(x) \equiv a_n\Phi_n(x)\}$ is the corresponding normalized sequence. The notations used are those of my monograph: *Théorie générale des polynomes orthogonaux de Tchebichef*, Mémorial des Sciences Mathématiques, Fasc. 66, 1934 (hereafter designated by M), to which the reader is referred for further details.

[2] Jacques Chokhate, *Sur une classe étendue de fractions continues algébriques et sur les polynomes de Tchebycheff correspondants*, Comptes Rendus, vol. 191 (1930), pp. 989–990.

1. We consider the sequence of OP

$$\Phi_n \equiv \Phi_n(x) \equiv \Phi_n(x; a, b; p), \tag{4}$$

that is,

$$\int_a^b p(x)\Phi_m(x)\Phi_n(x)dx = 0 \qquad (m \neq n;\, m, n = 0, 1, 2, \cdots).$$

Concerning $p(x)$, we make the following assumptions:

(i) $p(x)(x + a)^\sigma$ is finite as $x \to a$ (a finite, σ fixed). (Similarly for $x \to b$.) If (a, b) is infinite, say $b = \infty$, then $\lim_{x\to\infty} x^k p(x) = 0$ ($k = 0, 1, \cdots$). (Similarly if $a = -\infty$.)

(ii) $p(x)$ is of the following form:

$$p(x) = \frac{1}{A}\exp\left[\int \frac{B}{A}\, dx\right], \qquad \text{i.e., } A' + A\frac{p'}{p} = B. \tag{5}$$

(Naturally, we assume the existence of all integrals $\int_a^b x^n p(x)\, dx$ ($n = 0, 1, \cdots$).) Here $A \equiv A(x)$ and $B \equiv B(x)$ are certain fixed polynomials. To simplify our discussion, we assume that $A(x) > 0$, for $a < x < b$, i.e., $A(x)$ has no zeros inside (a, b). If A vanishes at $x = a$ (a finite), then it is of course assumed that B is such that (i) is satisfied. (Similarly for $x = b$.)

From the identity

$$\frac{1}{A}\exp\left[\int \frac{B}{A} dx\right] \equiv \frac{1}{AQ}\exp\left[\int \frac{BQ + AQ'}{AQ} dx\right]$$

it follows immediately that A, B in (5) may be replaced by $A_1 = AQ$, $B_1 = BQ + AQ'$, where $Q(x)$ is a fixed polynomial of the same nature as $A(x)$. We thus choose in (5) A so that

$$Ap = 0 \qquad \text{at } x = a, b, \tag{6}$$

where a, b are finite (by virtue of the first part of condition (i)). (For (a, b) infinite use the second part of the same condition.) Moreover, by virtue of (5),

$$(Ap)' = Bp. \tag{7}$$

In our discussion we shall also deal with the following sequence of OP:

$$u_n \equiv \Phi_n(x; a, b; p_1), \qquad p_1 = \Pi(x)p(x) \qquad (n = 0, 1, \cdots), \tag{8}$$

where $\Pi(x)$ is a fixed polynomial, of degree q, $\Pi(x) > 0$ in (a, b).

Here, by (5), we may take

$$A_1 = A\Pi, \qquad B_1 = B\Pi + 2A\Pi'. \tag{9}$$

A simple application of orthogonality gives (M, p. 26)

$$\Pi u_n = \sum_{i=n}^{n+q} h_i\Phi_i \qquad (h_i = \text{const.}). \tag{10}$$

Hereafter, Q, R and h, with various subscripts, stand respectively for polynomials or rational functions in x or constants, properly chosen and not necessarily the same in different formulas; $G_s(x) \equiv \sum_{i=0}^{s} g_i x^i$ stands for an arbitrary polynomial, of degree $\leqq s$; this degree will be generally denoted by δG_s.

We shall make constant use of the following lemma derived at once by a repeated application of (1).

LEMMA.

(i) $$\Phi_\mu(x) = Q_1\Phi_\nu(x) + Q_2\Phi_{\nu-1}(x) \qquad (\nu < \mu),$$

(ii) $$\sum_{i=\nu}^{\mu} h_i\Phi_i = Q_1\Phi_n + Q_2\Phi_{n-1} = Q_3\Phi_{n-1} + Q_4\Phi_{n-2} \qquad (\nu \leqq n \leqq \mu).$$

Moreover, in the "symmetric" case (i.e., all c_n in (1) vanish), Q_1, Q_2, $\cdots$ each contains only even or only odd powers of x (for so does each Φ_n).

It is seen that the coefficients of Q_1, Q_2, $\cdots$ involve the $c - s$ and $\lambda - s$ from (1). Note that in the symmetric case (10) becomes, if $\Pi(x) \equiv \Pi(-x)$,

(11) $$\Pi u_n = h_{n+q}\Phi_{n+q} + h_{n+q-2}\Phi_{n+q-2} + h_{n+q-4}\Phi_{n+q-4} + \cdots.$$

2. Write the orthogonality property of Φ_n as

(12) $$\int_a^b p\Phi_n G_{n-1}\,dx = 0 \qquad (n = 1, 2, \cdots).$$

(6) and (7) yield at once

(13) $$\begin{aligned} Ap\Phi_n G_\nu \Big|_a^b &= 0 = \int_a^b (Ap\Phi_n G_\nu)'\,dx \\ &= \int_a^b p[B\Phi_n + A\Phi_n']G_\nu\,dx + \int_a^b pA\Phi_n' G_\nu'\,dx. \end{aligned}$$

The relation (13) *is the basis of the subsequent discussion.*

Let

(14) $$\delta A = r, \qquad \delta B = s.$$

(13), through (12), leads to

(15) $$\int_a^b p[B\Phi_n + A\Phi_n']G_{n-r}\,dx = 0,$$

whence,

(16) $$B\Phi_n + A\Phi_n' = \sum_{i=n-r+1}^{n+\nu} h_i\Phi_i, \qquad \nu = \max(s, r-1).$$

It is important to note that *the summation on the right* (and similar summations in the subsequent discussion) *contains a fixed number of terms, independent of* n. By virtue of the lemma, we write

$$B\Phi_n + A\Phi_n' = Q_{1,n}\Phi_n + Q_{2,n}\Phi_{n-1}, \tag{16.1}$$

$$B\Phi_n + A\Phi_n' = Q_{3,n}\Phi_{n-1} + Q_{4,n}\Phi_{n-2}. \tag{16.2}$$

We get further from (16.1) by differentiation and combining with (16.2)

$$A^2\Phi_n'' + AQ_{5,n}\Phi_n' + AQ_{6,n}\Phi_n = Q_{7,n}\Phi_{n-1} + Q_{8,n}\Phi_{n-2}. \tag{17}$$

The desired differential equation for Φ_n is now obtained by eliminating Φ_{n-1}, Φ_{n-2} from (16.2), (17) and (1):

$$\begin{vmatrix} x - c_n & -\lambda_n & \Phi_n \\ Q_{3,n} & Q_{4,n} & A\Phi_n' + B\Phi_n \\ Q_{7,n} & Q_{8,n} & A^2\Phi_n'' + AQ_{5,n}\Phi_n' + AQ_{6,n}\Phi_n \end{vmatrix} = 0, \tag{18}$$

$$A^2Q_{9,n}\Phi_n'' + AQ_{10,n}\Phi_n' + Q_{11,n}\Phi_n = 0. \tag{19}$$

It is seen that $Q_{11,n}$ is divisible by A, since no Φ_n may have a common zero with A (the zeros of Φ_n lying between a and b), and the differential equation takes the desired form

$$AQ_n\Phi_n'' + B_n\Phi_n' + C_n\Phi_n = 0. \tag{20}$$

If B and A have a common factor C, write, in place of (16.1) and (16.2),

$$B_1\Phi_n + A_1\Phi_n' = Q_{1,n}\Phi_n + Q_{2,n}\Phi_{n-1}, \tag{21.1}$$

$$B_1\Phi_n + A_1\Phi_n' = Q_{3,n}\Phi_{n-1} + Q_{4,n}\Phi_{n-2}, \tag{21.2}$$

$$\left(B_1 = \frac{B}{C}, A_1 = \frac{A}{C}\right),$$

and proceed as above, with less computation; for the degrees of the polynomials involved will be reduced by $\delta C = r_1$. By hypothesis,

$$C > 0 \text{ inside } (a, b). \tag{22}$$

(15) now becomes

$$\int_a^b Cp[B_1\Phi_n + A_1\Phi_n']G_{n-r} = 0, \tag{23}$$

so that here

$$B_1\Phi_n + A_1\Phi_n' = \sum_{i=n-r+1}^{n+\nu_1} h_{i,n}' v_i \tag{24}$$

$$(v_i \equiv \Phi_i(x; a, b; pC); \ \nu_1 = \max(r - r_1 - 1, s - r_1)),$$

$$B_1\Phi_n + A_1\Phi_n' = Q_{5,n}v_n + Q_{6,n}v_{n-1}, \tag{25}$$

$$A_1\Phi_n'' + (B_1 + A_1')\Phi_n' + B_1'\Phi_n = Q_{5,n}'v_n + Q_{6,n}'v_{n-1} + Q_{5,n}v_n' + Q_{6,n}v_{n-1}', \tag{26}$$

and we may use the known properties of v_n, v_n' (see below).

3. The method just developed may be modified—not in principle, but in details—as follows. Take in (13) $\nu = n - \max(r, s + 1) \equiv n - \sigma - 1$, and we get

$$\int_a^b A p \Phi_n' G_{n-\sigma-1} = 0, \tag{27}$$

$$\begin{cases} A\Phi_n' = \sum_{i=n-\sigma}^{n+r-1} h_{i,n}\Phi_i , \\ A\Phi_n' = Q_{1,n}\Phi_n + Q_{2,n}\Phi_{n-1} = Q_{3,n}\Phi_{n-1} + Q_{4,n}\Phi_{n-2} . \end{cases} \tag{28}$$

We now proceed precisely as above (see (17), (18)), and we again obtain a differential equation for Φ_n of the type (20):

$$A\bar{Q}_n\Phi_n'' + \bar{B}_n\Phi_n' + \bar{C}_n\Phi_n = 0. \tag{29}$$

Comparing it with (20), we conclude that

$$\frac{\bar{Q}_n}{Q_n} = \frac{\bar{B}_n}{B_n} = \frac{\bar{C}_n}{C_n}, \tag{30}$$

for otherwise

$$\frac{\Phi_n'}{\Phi_n} = \frac{Q_n\bar{C}_n - \bar{Q}_n C_n}{Q_n\bar{B}_n - \bar{Q}_n B_n}.$$

This is impossible since the degree of Φ_n varies with n and Φ_n' has no factor in common with Φ_n, while the degrees of the polynomials on the right are fixed (by (28)). The relations (30) may lead to the explicit determination of the constants λ_n, c_n, $\cdots$ entering into (20) and (29).[3]

The case when $A = A_1C$ can be treated in the same manner as the above case: $A = A_1C$, $B = B_1C$.

4. The same method yields the solution of the following

Problem. Given the differential equation (3) for $\Phi_n(x;\, a,\, b;\, p)$. Find a similar equation for $u_n \equiv \Phi_n(x;\, a,\, b;\, p\Pi)$, as given in (8).

Solution. First, rewrite (10) as

$$\Pi u_n = Q_{1,n}\Phi_{n+1} + Q_{2,n}\Phi_n . \tag{31}$$

Secondly, reasoning as above and making use of (9), we get

$$A_1 p_1 u_n G_\nu \Big|_a^b = 0 = \int_a^b (A_1 p_1 u_n G_\nu')\, dx \quad (p_1(x) \equiv \Pi(x)p(x)),$$

[3] The following is a still simpler variation of the same method. We get (integrate by parts): $\int_a^b (Ap\phi_n')'G_{n-\nu}\, dx - \int_a^b (ApG_{n-\nu}')'\phi_n\, dx = 0$, whence, with $\nu = \max(r - 1, s)$:

$$A\Phi_n'' + B\Phi_n' = \sum_{i=n-\nu+1}^{n+\sigma} h_{i,n}'\Phi_i = Q_{n-1}\Phi_{n-1} + Q_{n-2}\Phi_{n-2} \qquad (\sigma = \max(r - 2, s - 1)).$$

The desired differential equation for Φ_n is obtained by eliminating Φ_{n-1}, Φ_{n-2} from (1), (28), and (*).

$$\int_a^b A_1 p_1 u_n' G_{n-\sigma_1-1}\, dx = \int_a^b A\Pi^2 p u_n' G_{n-\sigma_1-1}\, dx = 0$$

$$(\delta A_1 = r_1, \quad \delta B_1 = s_1, \quad \sigma_1 = \max\,(r_1 - 1, s_1)).$$

$$A\Pi^2 u_n' = \sum_{i=n-\sigma_1}^{n-1+r+2q} h_{i,n}' \Phi_i = Q_{3,n}\Phi_{n+1} + Q_{4,n}\Phi_n. \tag{32}$$

Combining with (31), we get

$$\begin{aligned} \Phi_{n+1} &= \Pi(R_{1,n}u_n + R_{2,n}u_n'), \\ \Phi_n &= \Pi(R_{3,n}u_n + R_{4,n}u_n'). \end{aligned} \tag{33}$$

Differentiating twice, substituting into (3) and clearing fractions, we obtain

$$Q_{5,n}u_n''' + Q_{6,n}u_n'' + Q_{7,n}u_n' + Q_{8,n}u_n = 0, \tag{34}$$

$$Q_{9,n}u_n''' + Q_{10,n}u_n'' + Q_{11,n}u_n' + Q_{12,n}u_n = 0. \tag{35}$$

Finally, combine (34) and (35), so as to eliminate u_n''', and the desired differential equation follows:

$$D_n u_n'' + E_n u_n' + F_n u_n = 0, \tag{36}$$

where D_n, E_n, F_n are polynomials in x of certain fixed degrees.

We shall not dwell here upon possible modifications and simplifications of the above procedure.

5. The general method, as exhibited in the foregoing, may be applied to the classical orthogonal polynomials in which case it yields very readily the classical differential equations, also the explicit expressions for the quantities $\lambda_n, c_n, S_n, \cdots$.

We now turn to two new cases.

(i) $\Phi_n(x; -\infty, \infty; e^{-\frac{1}{4}x^4})$, symmetric case. We take here $A = 1$, so that $B = -x^3$. The above considerations yield the following results.

$$(\alpha) \qquad \int_{-\infty}^{\infty} p\Phi_n' G_{n-4}\, dx = 0.$$

Proceeding as above, we obtain the desired differential equation in the determinantal form (18), which we shall not write down.

Here we make use of the following formula, valid in the symmetric case:[4]

$$d_{n,n-2} = -(\lambda_2 + \lambda_3 + \cdots + \lambda_n); \qquad \lambda_n = d_{n-1,n-3} - d_{n,n-2} \tag{37}$$

(compare coefficients in the recurrence relation (1)).

[4] The interest of this formula lies in the fact that if we know but the second highest coefficient of Φ_n, we can find λ_n, then the "normalizing coefficient" $a_n = (\lambda_1 \lambda_2 \cdots \lambda_n)^{-\frac{1}{2}}$, then Hankel's determinant $\Delta_n = (a_0 a_1 \cdots a_{n-1})^{-2}$ of order n formed by the moments $\alpha_i = \int_a^b p(x)x^i\, dx$ (M, p. 13). Illustration: the polynomials of Legendre and Hermite.

$$(\beta)\qquad \int_{-\infty}^{\infty} p(x)(-x^3\Phi_n + \Phi_n')G_n\,dx = 0; \qquad -x^3\Phi_n + \Phi_n' = -\Phi_{n+3} + h_n\Phi_{n+1}.$$

Comparing coefficients and making use of (37), we see that

$$h_n = \lambda_{n+1} + \lambda_{n+2} + \lambda_{n+3}.$$

Proceeding as above, we get a second differential equation in determinantal form which, if we expand and compare coefficients of x^{n+4}, x^{n+2}, yields

$$(38)\qquad 2(\lambda_2 + \lambda_3 + \cdots + \lambda_{n+1}) = -\lambda_{n+2}[1 - (\lambda_{n+1} + \lambda_{n+2})(\lambda_{n+2} + \lambda_{n+3})],$$

$$(39)\qquad \lambda_{n+2}(\lambda_{n+1} + \lambda_{n+2} + \lambda_{n+3}) = n + 1,$$

and the said equation takes the form

$$\begin{aligned} &M_n\Phi_n'' - x(x^2M_n + 2)\Phi_n' \\ (40)\qquad &+ \{x^2(nM_n - 2\lambda_{n+1}) - \lambda_{n+2}M_n[1 - (\lambda_{n+1} + \lambda_{n+2})(\lambda_{n+2} + \lambda_{n+3})]\}\Phi_n = 0; \\ &M_n = x^2 + \lambda_{n+1} + \lambda_{n+2}. \end{aligned}$$

Comparing this with the differential equation obtained in (α), we get the following relations for the λ's:

$$(41)\qquad 2(\lambda_2 + \lambda_3 + \cdots + \lambda_n) = n\lambda_n + \lambda_{n-1}\lambda_n\lambda_{n+1} = \lambda_{n+1}(n - 1 - \lambda_n),$$

$$(42)\qquad \lambda_{n+1}(\lambda_n + \lambda_{n+1} + \lambda_{n+2}) = n,$$

whence, $\lambda_{n+1} < n^{\frac12}$. We have further

$$\alpha_{2n-1} = 0; \qquad \alpha_{2n} = \int_{-\infty}^{\infty} e^{-\frac14 x^4}x^{2n}\,dx = 2^{n-\frac12}\Gamma[\tfrac14(2n+1)] \qquad (n \geqq 0);$$

$$\alpha_{2n} = (2n-3)\alpha_{2n-4} \quad (n \geqq 2); \qquad \alpha_0\alpha_2 = \pi 2^{\frac12};$$

$$\lambda_1 = \alpha_0, \qquad \lambda_2 = \frac{\alpha_2}{\alpha_0}, \qquad \lambda_3 = \frac{\alpha_0}{\alpha_2} - \frac{\alpha_2}{\alpha_0}, \qquad \cdots \quad [\text{M, pp. 9, 13}];$$

$$\Delta_1 = \alpha_0 = 2^{-\frac12}\Gamma(\tfrac14), \qquad \Delta_2 = \alpha_0\alpha_2, \qquad \Delta_3 = \alpha_2(\alpha_0^2 - \alpha_2^2), \qquad \cdots.$$

The above values for λ_1, λ_2, λ_3, combined with (42), enable us to compute λ_n (the only parameter entering into the differential equation (40)) for any given n.[5]

(ii) $\Phi_n(x; -1, 1; (1 - x^2)^{\frac12}(1 - \mu x^2)^{-1})$, $\mu \leqq 1$, again a symmetric case. We take here $A = (1 - x^2)(1 - \mu x^2)$, so that $B = -3x(1 - \mu x^2)$. Thus, A and B have a common factor whose presence introduces marked improvements in our general considerations, as we proceed to show.

We have here

$$(43)\qquad \int_{-1}^{1} Ap\Phi_n'G_{n-4}\,dx = \int_{-1}^{1} (1 - x^2)^{\frac12}\Phi_n'G_{n-4}\,dx = 0.$$

[5] It would be of interest to devise an explicit expression for Δ_n as a function of n. This would yield an expression for λ_n. The positiveness of all Δ_n yields inequalities for $\Gamma(\frac14)$. Thus, from $\Delta_4 > 0$, $\Delta_5 > 0$,

$$3^{\frac14} > \Gamma(\tfrac14)\pi^{-\frac12}2^{-\frac34} > 2^{\frac14}.$$

Introduce the following sequences of Jacobi polynomials

$$u_n \equiv J_n(x; \tfrac{5}{2}, \tfrac{5}{2}), \qquad v_n \equiv J_n(x; \tfrac{3}{2}, \tfrac{3}{2}) \tag{44}$$

for which, as is known (M, pp. 17, 33),

$$\begin{cases} (1 - x^2)v_n'' - 3xv_n' + n(n + 2)v_n = 0, \\ v_n(\cos \varphi) = \dfrac{\sin (n + 1)\varphi}{2^n \sin \varphi}, \qquad \lambda_n = \tfrac{1}{4} \; (n \geqq 2), \\ v_n' \equiv nu_{n-1}. \end{cases} \tag{45}$$

(43) yields at once

$$\Phi_n' = nu_{n-1} + (n - 2)h_n u_{n-3} = v_n' + h_n v_{n-2}', \tag{46}$$

$$\Phi_n = x^n + d_{n,n-2}x^{n-2} + \cdots = v_n + h_n v_{n-2} + h_{1,n}. \tag{47}$$

Here $h_{1,n} = 0$. For $n = 2m + 1$, this follows immediately from $\Phi_{2m+1}(0) = v_{2m+1}(0) = 0$ $(m = 0, 1, 2, \cdots)$. If n is even, we make use of the recurrence relation (1) for Φ_{2n}, v_n.

We get further, by (46), (45), (23),

$$(1 - x^2)\Phi_n'' - 3x\Phi_n' = -n(n + 2)v_n - h_n(n - 2)nv_{n-2}. \tag{48}$$

We have also (by (23))

$$(1 - x^2)\Phi_n' - 3x\Phi_n = -(n + 3)v_{n+1} + h_{2,n}v_{n-1} + h_{3,n}\lambda_{n-1}'v_{n-3}$$

(λ_n' corresponds to v_n),

and this relation, through (1), leads to

$$\begin{aligned} x(1 - x^2)\Phi_n' - 3x^2\Phi_n = {} & [-(n + 3)x^2 - h_{3,n} - (n + 3)\lambda_n' + h_{2,n}]v_n \\ & + [\lambda_n'(h_{3,n}x + (n + 3)\lambda_{n+1}') + h_{2,n}]v_{n-2}. \end{aligned} \tag{49}$$

The desired differential equation is now obtained by eliminating v_n, v_{n-2} from (47), (48), (49).

It remains to determine the various constants in the above relations. By comparing coefficients in (47), (49) and making use of

$$v_n = x^n - \frac{n - 1}{4}x^{n-2} + \frac{(n - 2)(n - 3)}{32}x^{n-4} \mp \cdots,$$

we get

$$\begin{gathered} h_n = d_{n,n-2} + \frac{n - 1}{4} = \frac{n - 1}{4} - (\lambda_2 + \cdots + \lambda_n), \\ d_{n,n-4} = \cdots, \qquad h_{2,n} = \cdots, \qquad \cdots. \end{gathered} \tag{50}$$

All reduces to finding h_n. For this purpose we make use of

$$0 = \int_{-1}^{1} \frac{(1 - x^2)^{1/2}}{1 - \mu x^2} x^{\epsilon} \Phi_n(x)\, dx = \int_{-1}^{1} \frac{(1 - x^2)^{1/2}}{1 - \mu x^2} x^{\epsilon}(v_n + h_n v_{n-2})\, dx = 0$$

$$(\epsilon = 0, 1 \text{ for } n \text{ even, odd respectively}),$$

$$\int_0^{\pi} \frac{\cos ax\, dx}{1 + p \cos x} = \frac{\pi}{(1 - p^2)^{1/2}} \left\{ \frac{(1 - p^2)^{1/2} - 1}{p} \right\}^a \qquad (|p| < 1),$$

where a is a positive integer or 0. We get[6]

$$h_n = -\frac{2\nu + 2 - \mu}{4\mu} \qquad (\nu = (1 - \mu)^{1/2}), \qquad d_{n,n-2} = \cdots, \qquad \cdots$$

$$(1 - x^2)\left[1 + \frac{\nu}{n} - \mu x^2\right]\Phi_n'' + \left\{\mu x^2 - \left[2(1 - \mu) + \frac{3\nu}{n} + 1\right]\right\} x\Phi_n' + \{(n + \nu)(n - 2\nu) - n^2 \mu x^2\} \Phi_n = 0. \tag{51}$$

Note that our present method gives not only the differential equation under discussion, but also (and without any further considerations) the explicit expressions for $\Phi_n(x)$, also for λ_n, $d_{n,n-2}$, $\cdots$.

The corresponding formulas for the limiting cases $p(x) = (1 - x^2)^{-1/2}$, $(1 - x^2)^{1/2}$ are furnished by the same formulas, if we let $\mu \to 1$, 0 respectively; in the latter case we take $\nu = -(1 - \mu)^{1/2}$, since here $h_n = 0$ for all n.

Remark. Using the known transformation from the symmetric to the corresponding non-symmetric case (M, p. 19), we readily derive from the foregoing results differential equations of the desired type for

$$\Phi_n(x; 0, \infty; e^{-x^2} x^{\sigma}), \qquad \Phi_n\left(x; 0, 1; \frac{(1 - x)^{1/2}}{1 - \mu x} x^{\sigma}\right), \qquad \sigma = \pm\tfrac{1}{2}. \tag{52}$$

6. We now return to the general discussion. Take

$$A = (x - a)(b - x), \tag{53}$$

where (a, b) is the interval of orthogonality (with the customary agreement to replace $x - a$ by unity, if $a = -\infty$). (Similarly for $b - x$.) Make use of the following simple identity (integrate by parts):

$$\int_a^b f[\psi\varphi']'\, dx = \int_a^b \varphi[\psi f']'\, dx, \qquad \psi(a) = \psi(b) = 0, \tag{54}$$

[6] The explicit expression for $\Phi_n(x)$ may be obtained by the method of G. Szegö: *Ueber die Entwickelung einer willkürlichen Funktion nach den Polynomen eines Orthogonalsystems*, Math. Zeitschrift, vol. 12(1922), pp. 61–94. Cf. also S. Bernstein, Jour. des Math., (9), vol. 9(1930), p. 175 and Comm. Soc. Math. Kharkoff, 1930. J. Geronimus (Kharkoff, Russia) obtained and communicated to the author this differential equation without, however, revealing the method used.

which is valid for any two functions f, φ, each possessing in (a, b) first and second derivatives. We get, by virtue of (6), (7) and the orthogonality property of φ_n,

$$\int_a^b (Ap\varphi_n')G_{n-1}\,dx = \int_a^b \varphi_n(ApG_{n-1}')'\,dx = \int_a^b p\varphi_n BG_{n-1}'\,dx. \tag{55}$$

Consider, first, the classical OP of Jacobi (J) in $(-1, 1)$, Laguerre (L), Hermite (H). Here $\delta B = 1$; hence,

$$\int_a^b (Ap\varphi_n')'G_{n-1}'\,dx = 0,$$

$$(Ap\varphi_n')' = -C_n p\varphi_n \qquad (C_n = \text{const.}). \tag{56}$$

This is the classical differential equation (M, p. 33), with

$$C_n = n(n + \alpha + \beta - 1)\ (\text{J}), \qquad n\ (\text{L}), \qquad 2n\ (\text{H}). \tag{57}$$

The consequences which follow directly from (56) have not been exhausted yet, as we proceed to show. In the first place, we write

$$\begin{cases} Ap\varphi_n' \Big|_a^x = A(x)p(x)\varphi_n'(x) = -C_n \displaystyle\int_a^x p(t)\varphi_n(t)\,dt, \\ A(x)p(x)\varphi_n'(x) = O(C_n), \ = O(n^2)\ (\text{J}), \ = O(n)\ (\text{L, H}) \quad (a \leqq x \leqq b) \end{cases} \tag{58}$$

since, by the Schwarz inequality, each of

$$\int_a^x p(t)\varphi_n(t)\,dt, \qquad \int_a^x p(t)\varphi_m(t)\varphi_n(t)\,dt = O(1) \qquad (a \leqq x \leqq b).$$

Making use of the relation[7]

$$\varphi_n'(x; p) \equiv C_n^{1/2}\varphi_{n-1}(x; Ap), \tag{59}$$

we get from (58)

$$A(x)p(x)\varphi_{n-1}(x; Ap) = -C_n^{1/2}\int_a^x p(t)\varphi_n(t)\,dt = O(C_n^{1/2}) \qquad (a \leqq x \leqq b), \tag{60}$$

$$A(x)p(x)\varphi_{n-1}(x; Ap) = O(n)\ (\text{J}), \quad O(n^{1/2})\ (\text{H, L}). \tag{61}$$

On the other hand, if we make use of the known asymptotic expression for $\varphi_n(x)$ (M, pp. 62–64), we get from (60)

$$\begin{aligned} &\int_{-1}^x p(t)\varphi_n(t)\,dt = O(n^{-1}) && (\text{J}, -1 + \epsilon \leqq x \leqq 1 - \epsilon), \\ &\int_a^x p(t)\varphi_n(t)\,dt = O(n^{-1/4}) && (\text{L}, -\epsilon \leqq x \leqq C; \text{H}, -C \leqq x \leqq C). \end{aligned} \tag{62}$$

[7] J. Shohat, *On the development of functions in series of orthogonal polynomials*, Bull. Amer. Math. Soc., vol. 41 (1935), pp. 49–82; p. 75.

(Here and hereafter ϵ and C stand respectively for an arbitrarily small and an arbitrarily large, but fixed, positive quantity.) Another result, derived very simply from (56), through (60), concerns the expansion of functions in series of classical OP.

It is known that for these OP Parseval's Formula holds, so that

$$(63)\qquad \int_a^x p(t)f(t)\,dt = \sum_{n=0}^{\infty} f_n \int_a^x p(t)\varphi_n(t)\,dt, \qquad f_n = \int_a^b p(t)f(t)\varphi_n(t)\,dt$$

(if $\int_a^b p(t)f^2(t)\,dt$ exists; convergence uniform, $a \leqq x \leqq b$). Hence, by (60),

$$(64)\qquad F(x) \equiv \int_a^x p(t)f(t)\,dt = \sum_{n=1}^{\infty} - f_n C_n^{-\frac12} A(x)p(x)\varphi_{n-1}(x; Ap)$$

(convergence uniform, $a \leqq x \leqq b$), whence we get

THEOREM I. *If $F(x)$ may be represented as $\int_a^x p(t)f(t)\,dt$ and $\int_a^b p(t)f^2(t)\,dt$ exists, then $F(x)/A(x)p(x)$ can be expanded in a series of the polynomials $\{\varphi_n(x; Ap)\}$, namely,*

$$(65)\qquad \frac{F(x)}{A(x)p(x)} = \sum_{n=1}^{\infty} - C_n^{-\frac12}\left(\int_a^b p(t)f(t)\varphi_n(t)\,dt\right)\varphi_{n-1}(x; Ap),$$

which converges uniformly in any interval wholly inside (a, b).[8]

Note that no use was made here of the asymptotic properties of $\varphi_n(x)$ and that we obtained explicit formulas for the coefficients in the expansion (65) in terms of those for $f(x)$. Note also that, without any further assumption concerning $f(x)$, we get for the remainder in the expansion (63), by (62),

$$(66)\qquad \begin{aligned} &\left| \int_a^x p(t)f(t)\,dt - \sum_{i=0}^{n} f_i \int_a^x p(t)\varphi_i(t)\,dt \right| \\ &\leqq \left[\sum_{i=n+1}^{\infty} f_i^2 \cdot \sum_{i=n+1}^{\infty} \left(\int_a^x p(t)\varphi_i(t)\,dt \right)^2 \right]^{\frac12} = o(n^{-\frac12})\ (\mathrm{J}),\ o(n^{-\frac12})\ (\mathrm{H, L}), \end{aligned}$$

where x is as in (62). With $p(x) = 1$ and (a, b) finite, this leads to the approximation of a certain class of absolutely continuous functions, by means of polynomials—certain indefinite integrals of Legendre polynomials.

Consider, next, the case where $\delta B \leqq 2$:

$$(67)\qquad B = hx^2 + kx + l.$$

While the differential equation for $\varphi_n(x)$ may be derived by the method developed above, here we center our attention on $\varphi_n'(x)$ and on

$$(68)\qquad K_n(x) \equiv \sum_{i=0}^{n} \varphi_i^2(x) = \lambda_{n+2}^{\frac12}[\varphi_{n+1}'(x)\varphi_n(x) - \varphi_n'(x)\varphi_{n+1}(x)] \quad (\mathrm{M, p.\ 25}).$$

[8] Cf. loc. cit. in footnote 7, pp. 71–76; see also (for Laguerre and Hermite polynomials) H. Weyl, *Singuläre Integralgleichungen* ⋯ , Thesis, Göttingen (1908), pp. 63, 74, where $F(x)$ is subject to more restrictive conditions.

By (5),

$$p(x) = e^{-hx}(x-a)^{\alpha-1}(b-x)^{\beta-1} \qquad (\alpha, \beta > 0;\ a, b \text{ finite}), \tag{69}$$

$$p(x) = e^{\frac{1}{2}hx^2+rx}(x-a)^{\alpha-1} \qquad (a \text{ finite},\ b = \infty;\ r < 0, \text{ if } h = 0). \tag{70}$$

[In case $(a, b) = (-\infty, \infty)$, we get again Hermite polynomials.] Apply (55) to $G_{n-2}(x)$, and we get

$$\int_a^b (Ap\varphi_n')' G_{n-2}\,dx = \int_a^b p\varphi_n(BG_{n-2}')\,dx = 0,$$

$$(Ap\varphi_n')' + C_n\varphi_n = h_{n,n+1}\varphi_{n+1} + h_{n,n-1}\varphi_{n-1}, \tag{71}$$

where C_n, $h_{n,n+1}$, $h_{n,n-1}$ are constants. Comparing the coefficients of x^n gives at once

$$h_{n,n+1} = nh\cdot\frac{a_n}{a_{n+1}} = nh\lambda_{n+2}^{\frac{1}{2}}.$$

Moreover, by the orthogonality property and making use of (55), we have

$$h_{n,n-1} = \int_a^b (Ap\varphi_n')'\varphi_{n-1}\,dx = \int_a^b pB\varphi_n\varphi_{n-1}'\,dx = (n-1)h\lambda_{n+1}^{\frac{1}{2}},$$

whence,

$$A\varphi_n'' + B\varphi_n' + C_n\varphi_n = h[n\lambda_{n+2}^{\frac{1}{2}}\varphi_{n+1} + (n-1)\lambda_{n+1}^{\frac{1}{2}}\varphi_{n-1}], \tag{72}$$

$$\begin{cases} C_n = n(n-1) - nk - hS_n - nhc_{n+1}, & \text{in case (69)}, \\ C_n = -nk - hS_n - nhc_{n+1}, & \text{in case (70)}. \end{cases} \tag{73}$$

Making use of the recurrence relation (M, p. 24), we obtain

$$\lambda_{n+2}^{\frac{1}{2}}\varphi_{n+1}(x) = (x - c_{n+1})\varphi_n(x) - \lambda_{n+1}^{\frac{1}{2}}\varphi_{n-1}(x).$$

We get further

$$A\varphi_n'' + B\varphi_n' + D_n\varphi_n = -h\lambda_{n+1}^{\frac{1}{2}}\varphi_{n-1}, \tag{74}$$

$$D_n \equiv C_n - nh(x - c_{n+1}) = \begin{cases} n(n-1) - hS_n - n(hx + k), & \text{in case (69)}, \\ -hS_n - n(hx + k), & \text{in case (70)}. \end{cases} \tag{75}$$

Rewrite (74), by virtue of (5), (7), as

$$(Ap\varphi_n')' = -h\lambda_{n+1}^{\frac{1}{2}}p\varphi_{n-1} - D_n p\varphi_n,$$

whence,

$$A(x)p(x)\varphi_n'(x) = -h\lambda_{n+1}^{\frac{1}{2}}\int_a^x p(t)\varphi_{n-1}(t)\,dt - \int_a^x p(t)D_n(t)\varphi_n(t)\,dt. \tag{76}$$

We now turn to $K_n(x)$. By virtue of (68), we readily get from (74) the following differential equation for $K_n(x)$:

$$
(77)\qquad \begin{aligned} AK_n'(x) + BK_n(x) &= \lambda_{n+2}^{1/2}(D_n - D_{n+1})\varphi_n\varphi_{n+1} \\ &\qquad + h\lambda_{n+2}^{1/2}[\lambda_{n+2}^{1/2}\varphi_n^2 - \lambda_{n+1}^{1/2}\varphi_{n+1}\varphi_{n-1}], \end{aligned}
$$

whence, by (75),[9]

$$
(78)\qquad \begin{aligned} A(x)p(x)K_n(x) &= \lambda_{n+2}^{1/2}\int_a^x [D_n(t) - D_{n+1}(t)]\varphi_n(t)\varphi_{n+1}(t)\,dt \\ &\qquad + h\lambda_{n+2}^{1/2}\int_a^x [\lambda_{n+2}^{1/2}\varphi_n^2(t) - \lambda_{n+1}^{1/2}\varphi_{n+1}(t)\varphi_{n-1}(t)]\,dt, \end{aligned}
$$

$$
(79)\qquad \begin{aligned} D_n(x) - D_{n+1}(x) &= -2n + hc_{n+1} + hx + k,\ hc_{n+1} + hx + k \\ &\qquad \text{(in cases (69), (70) respectively).} \end{aligned}
$$

In case (a, b) is finite, reduced, without loss of generality, to $(-1, 1)$, we know that

$$\lambda_n = O(1), \qquad c_n = O(1), \qquad S_n = O(n).$$

Hence, here

$$(80)\qquad C_n = n(n-1)[1 + O(n^{-1})];$$

$$(81)\qquad A(x)p(x)K_n(x) = O(n) \qquad (-1 \leqq x \leqq 1);$$

$$
(82)\qquad \begin{aligned} [A(x)p(x)]^{1/2}\,|\,\varphi_n(x)\,| &= O(n^{1/2}) \qquad (-1 \leqq x \leqq 1), \\ |\,\varphi_n(x)\,| &= O(n^{1/2}) \qquad (-1 + \epsilon \leqq x \leqq 1 - \epsilon); \end{aligned}
$$

and, by the Markoff-Bernstein Theorem,

$$
(83)\qquad \begin{aligned} &|\,\varphi_n'(x)\,| = O(n^{3/2}) \\ &\qquad (-1 + \epsilon + \epsilon' \leqq x \leqq 1 - \epsilon - \epsilon', \text{ i.e., } x \text{ inside } (-1, 1)). \end{aligned}
$$

The estimate (81) may be applied to the study of the expansion

$$
\begin{aligned} f(x) &\sim \sum_{n=0}^{\infty}\left[\int_{-1}^{1} pf\varphi_n\,dx\right]\varphi_n(x) \equiv \sum_{n=0}^{\infty} f_n\varphi_n(x) \\ &= \sum_{i=0}^{n} f_i\varphi_i(x) + R_n(x) \equiv S_n + R_n, \end{aligned}
$$

where $f(x)$ is assumed continuous in $(-1, 1)$. We have, denoting by $E_n(f)$ the "best approximation"—in the Tchebycheff sense—of $f(x)$ on $(-1, 1)$ by polynomials of degree $\leqq n$,[10]

[9] Cf., for the classical orthogonal polynomials, J. Shohat and C. Winston, *On mechanical quadratures*, Rendic. Circ. Mat. Palermo, vol. 58(1934), pp. 1–13; pp. 4–5.

[10] Loc. cit. (footnote 7), p. 59.

$$(84)\qquad |R_n(x)| \leqq E_n(f)\{1 + [K_n(x)]^{\frac{1}{2}}\} = O(E_n(f)n^{\frac{1}{2}}) \qquad (-1 + \epsilon \leqq x \leqq 1 - \epsilon)$$

by (81).

This leads to

THEOREM II. *The expansion of $f(x)$ in a series of* OP, *where* $p(x) = e^{-hx}(1 + x)^{\alpha-1}(1 - x)^{\beta-1}$ $(\alpha, \beta > 0)$, *converges uniformly in any interval wholly inside* $(-1, 1)$, *provided $f(x)$ satisfies therein a Lipschitz condition of order* $> \frac{1}{2}$.

Some of the foregoing results can be, and have been, improved, by use of more refined methods. Here they have been derived by means of very simple considerations as a direct sequel to the differential equation for the OP under discussion.

The same elementary considerations lead to new and important results in the theory of mechanical quadratures, as we proceed to show in the closing section.

7. We turn once more to the classical OP, i.e., we let $h = 0$ in (67). Formula (78) now becomes

$$(85)\qquad A(x)p(x)K_n(x) = \lambda_{n+2}^{\frac{1}{2}}(k - 2n) \int_a^x p(t)\varphi_n(t)\varphi_{n+1}(t)\,dt.$$

Denoting the zeros of $\varphi_n(x)$ by $x_{i,n}$ $(1 \leqq i \leqq n)$, with

$$(86)\qquad a < x_{1,n+1} < x_{1,n} < x_{2,n+1} < \cdots < x_{n,n} < x_{n+1,n+1} < b,$$

we learn from (85) that

$$(87)\qquad A(x)p(x)K_n(x) \begin{cases} \text{is maximum at } x = x_{i,n+1} & (1 \leqq i \leqq n + 1), \\ \text{is minimum at } x = x_{j,n} & (1 \leqq j \leqq n); \end{cases}$$

$$(88)\qquad \begin{aligned} &p_1(x_{1,n+1})K_n(x_{1,n+1}) > p_1(x_{1,n})K_n(x_{1,n}) < p_1(x_{2,n+1})K_n(x_{2,n+1}) > \cdots \\ &\qquad > p_1(x_{n,n})K_n(x_{n,n}) < p_1(x_{n+1,n+1})K_n(x_{n+1,n+1}) \qquad (p_1 \equiv Ap). \end{aligned}$$

Hence,

$$(89)\qquad p_1(x_{i,n})K_n(x_{i,n}) < p_1(x_{i+\sigma,n+1})K_n(x_{i+\sigma,n+1}) \qquad (1 \leqq i \leqq n;\ \sigma = 0, 1).$$

Introduce the Gaussian Mechanical Quadratures Formula

$$(90)\qquad \int_a^b p(x)f(x)\,dx \cong \sum_{i=1}^{n} H_{i,n}f(x_{i,n}), \qquad H_{i,n} = \frac{1}{K_n(x_{i,n})} = \frac{1}{K_{n-1}(x_{i,n})}.$$

Then (88), (89) become

$$(91)\qquad \begin{aligned} &\frac{H_{1,n+1}}{p_1(x_{1,n+1})} < \frac{H_{1,n}}{p_1(x_{1,n})} > \frac{H_{2,n+1}}{p_1(x_{2,n+1})} < \cdots < \frac{H_{n,n}}{p_1(x_{n,n})} > \frac{H_{n+1,n+1}}{p_1(x_{n+1,n+1})}, \\ &\qquad\qquad \frac{H_{i,n}}{p_1(x_{i,n})} < \frac{H_{i+\sigma,n+1}}{p_1(x_{i+\sigma,n+1})} \qquad (1 \leqq i \leqq n;\ \sigma = 0, 1). \end{aligned}$$

We have further

(92) $p_1'(x) = Bp; \qquad B = \alpha - \beta - (\alpha + \beta)x \text{ (J)}, \ \alpha - x \text{ (L)}, \ -2x \text{ (H)}.$

Hence, $p_1(x)$ is increasing in (a, c) and decreasing in (c, b), where

(93) $$c = \frac{\alpha - \beta}{\alpha + \beta} \text{ (J)}, \ \alpha \text{ (L)}, \ 0 \text{ (H)}; \quad c = x_{1,1}.$$

(Take $n = 1$ in the differential equation for the classical OP.) (91), combined with (86), now gives[11]

(94) $$H_{i,n} > \frac{p_1(x_{i,n})}{p_1(x_{i,n+1})} H_{i,n+1}; \qquad H_{i,n} > \frac{p_1(x_{i,n})}{p_1(x_{i+1,n+1})} H_{i+1,n+1}.$$

Consider, first, Jacobi polynomials. We learn from (91) that

(95) $$\begin{gathered} \frac{H_{1,n}}{p_1(x_{1,n})} > \frac{H_{1,n+1}}{p_1(x_{1,n+1})} > \frac{H_{1,n+2}}{p_1(x_{1,n+2})} > \cdots, \\ \frac{H_{n,n}}{p_1(x_{n,n})} > \frac{H_{n+1,n+1}}{p_1(x_{n+1,n+1})} > \frac{H_{n+2,n+2}}{p_1(x_{n+2,n+2})} > \cdots; \end{gathered}$$

and this assures the existence of the following limits:

(96) $$\lim_{n\to\infty} \frac{H_{1,n}}{p_1(x_{1,n})} = h' \ (\geqq 0), \qquad \lim_{n\to\infty} \frac{H_{n,n}}{p_1(x_{n,n})} = h'' \ (\geqq 0).$$

We proceed to show that *both h' and h'' are positive.* For this purpose we make use of Tchebycheff inequalities

$$H_{1,n} > \int_a^{x_{1,n}} p(x)\, dx, \qquad H_{n,n} > \int_{x_{n,n}}^b p(x)\, dx.$$

We get

$$\frac{H_{1,n}}{p_1(x_{1,n})} > \frac{\int_{-1}^{x_{1,n}} (1 + x)^{\alpha-1}(1 - x)^{\beta-1}\, dx}{(1 + x_{1,n})^{\alpha}(1 - x_{1,n})^{\beta}} > \frac{1}{\alpha}\cdot\frac{1}{1 - x_{1,n}} > \frac{1}{2\alpha}.$$

Similarly,

$$\frac{H_{n,n}}{p_1(x_{n,n})} > \frac{1}{2\beta}.$$

(A still simpler procedure is to employ the almost evident inequalities

$$H_{1,n} = \frac{1}{K_n(x_{1,n})} > \frac{1}{K_n(-1)}, \qquad H_{n,n} = \frac{1}{K_n(x_{n,n})} > \frac{1}{K_n(1)}$$

[11] This is better than $H_{i,n} > H_{i,n+1}$, $H_{i+1,n+1}$, respectively, as given in page 2 of the reference in footnote 9 and yields important results below.

and to make use of the known values of $K_n(\pm 1)$.) Still easier is it to obtain upper bounds for $H_{1,n}$, $H_{n,n}$. By (91),

$$\frac{H_{1,1}}{p_1(x_{1,1})} > \frac{H_{2,2}}{p_1(x_{2,2})} > \cdots > \frac{H_{n,n}}{p_1(x_{n,n})} > \cdots,$$

$$\frac{H_{1,1}}{p_1(x_{1,1})} > \frac{H_{1,2}}{p_1(x_{1,2})} > \cdots > \frac{H_{1,n}}{p_1(x_{1,n})} > \cdots.$$

Here we use

$$H_{1,1} = \int_a^b p(x) \frac{\varphi_1(x)}{(x - x_{1,n})\phi_1'(x)}\,dx = \int_a^b p(x)\,dx,$$

and we get[12]

$$h',\, h'' < \frac{(\alpha+\beta)^{\alpha+\beta}\Gamma(\alpha)\Gamma(\beta)}{2\alpha^\alpha\cdot\beta^\beta\Gamma(\alpha+\beta)}. \tag{97}$$

Similar simple considerations, applied to Laguerre polynomials (where we may use the inequality $H_{1,n} > 1/K_n(0)$), yield

$$\frac{1}{\alpha} < \frac{H_{1,n}}{p_1(x_{1,n})} < \frac{\Gamma(\alpha)}{e^{-\alpha}\alpha^\alpha}.$$

(Note the resulting inequality, valid for any $\alpha > 0 : \Gamma(\alpha) > e^{-\alpha}\alpha^{\alpha-1}$.) We summarize our results as follows. *For Jacobi polynomials* [*in* $(-1, 1)$] $H_{1,n}$ *and* $H_{n,n}$ *behave asymptotically* ($n \to \infty$) *like* $(1 + x_{1,n})^\alpha$ *and* $(1 - x_{n,n})^\beta$ *respectively, or, which is the same, like* $\int_{-1}^{x_{1,n}} p(x)\,dx$, $\int_{x_{n,n}}^1 p(x)\,dx$ *respectively. For Laguerre polynomials,* $H_{1,n}$ *behaves asymptotically like* $x_{1,n}^\alpha$, *or, which is the same, like* $\int_0^{x_{1,n}} p(x)\,dx$. *Namely,*

$$\lim_{n\to\infty} \frac{H_{1,n}}{(1 + x_{1,n})^\alpha} = J'(\alpha, \beta), \qquad \lim_{n\to\infty} \frac{H_{n,n}}{(1 + x_{n,n})^\beta} = J''(\alpha, \beta),$$

$$\frac{2^{\beta-1}(\alpha+\beta)^{\alpha+\beta}\Gamma(\alpha)\Gamma(\beta)}{\alpha^\alpha\beta^\beta\Gamma(\alpha+\beta)} > J'(\alpha,\beta) \geqq \frac{2^{\beta-1}}{\alpha}, \tag{98, J}$$

$$\frac{2^{\alpha-1}(\alpha+\beta)^{\alpha+\beta}\Gamma(\alpha)\Gamma(\beta)}{\alpha^\alpha\beta^\beta\Gamma(\alpha+\beta)} > J''(\alpha,\beta) \geqq \frac{2^{\alpha-1}}{\beta}.$$

$$\lim_{n\to\infty} \frac{H_{1,n}}{\int_{-1}^{x_{1,n}} p(x)\,dx} \text{ exists and is } \geqq 1; \qquad \lim_{n\to\infty} \frac{H_{n,n}}{\int_{x_{n,n}}^{1} p(x)\,dx} \text{ exists and is } \geqq 1.$$

[12] As a concrete illustration, we may utilize the trigonometric polynomials ($\alpha = \beta = \frac{1}{2}$). Here $H_{1,n} = H_{2,n} = \cdots = H_{n,n} = \pi n^{-1}$, $p_1(x_{1,n}) = p_1(x_{n,n}) = (1 - x_{1,n}^2)^{\frac{1}{2}} = \pi(2n)^{-1}[1 + O(n^{-2})]$, so that $\lim_{n\to\infty} H_{1,n}/p_1(x_{1,n}) = 2$.

$$\Gamma(\alpha)e^{\alpha}\alpha^{-\alpha} > L'(\alpha) \equiv \lim_{n\to\infty} \frac{H_{1,n}}{x_{1,n}^{\alpha}} \geqq \frac{1}{\alpha};$$

$$\text{(98, L)} \qquad \lim_{n\to\infty} \frac{H_{1,n}}{\int_0^{x_{1,n}} p(x)\,dx} \text{ exists and is } \geqq 1.$$

These formulas give the asymptotic expressions for $H_{1,n}$, $H_{n,n}$ if those for $x_{1,n}$, $x_{n,n}$ are known, and vice versa (see below). The known relation of Hermite polynomials to the polynomials of Laguerre, with $\alpha = \frac{1}{2}, \frac{3}{2}$ (M, p. 20), gives at once, with obvious notations,

$$H_{n+1,2n}\,(\mathrm{H}) = \frac{H_{1,n}}{2}\,(\mathrm{L}; \alpha = \tfrac{1}{2}); \qquad H_{n+2,2n+1}\,(\mathrm{H}) = \frac{H_{1,n}}{2x_{1,n}}\,(\mathrm{L}; \alpha = \tfrac{3}{2}),$$

and by the above,

$$\lim_{n\to\infty} \frac{H_{n+1,2n}}{x_{n+1,2n}} = H', \qquad (\tfrac{1}{2}\pi e)^{\frac{1}{2}} > H' \geqq 1;$$

$$\text{(98, H)}$$

$$\lim_{n\to\infty} \frac{H_{n+2,2n+1}}{x_{n+2,2n+1}} = H'', \qquad (\tfrac{1}{27}\pi e^3)^{\frac{1}{2}} > H'' \geqq \tfrac{1}{3}.$$

Thus, for Hermite polynomials, the coefficient in the mechanical quadrature formula corresponding to the zero nearest to 0 behaves asymptotically like this zero itself, which is in accordance with a result obtained by Winston[13] and derived by a less elementary method (Sonine's method).

On the other hand, if we make use of the asymptotic expression for some of the $H_{i,n}$ in Hermite's case,[14] we get from the above formulas, without any further consideration, not only the true order, but also the asymptotic expression $(n \to \infty)$ for the corresponding zeros. Thus,

$$\text{(99)} \quad \begin{cases} (\mathrm{H}) \text{ exists } \lim\limits_{n\to\infty} n^{\frac{1}{2}} x_{n+1,2n} = \dfrac{\pi}{2H'}; & \text{exists } \lim\limits_{n\to\infty} n^{\frac{1}{2}} x_{n+2,2n+1} = \dfrac{\pi}{2H''}; \\ (\mathrm{L}) \text{ exists } \lim\limits_{n\to\infty} n x_{1,n} = \dfrac{\pi^2}{4H'^2}\,(\alpha = \tfrac{1}{2}); & \text{exists } \lim\limits_{n\to\infty} n x_{1,n} = \dfrac{\pi^2}{4H''^2}\,(\alpha = \tfrac{3}{2}). \end{cases}$$

[The existing estimates generally deal with the largest zeros for (H), (L).] These asymptotic estimates may be combined with the estimate of the minimum distance $\delta(n)$ between two successive zeros for Hermite polynomials. In fact, according to Hille,[15] $2x_{n+1,2n}\,(\mathrm{H}) = \delta(2n)$; $x_{n+2,2n+1}\,(\mathrm{H}) = \delta(2n+1)$.

The author hopes to return to the considerations developed above.

THE UNIVERSITY OF PENNSYLVANIA.

[13] C. Winston, *On mechanical quadratures formulae* $\cdots$, Annals of Math., vol. 35(1934), pp. 658–677.

[14] Loc. cit. (footnote 7), p. 13; also loc. cit. (footnote 9), p. 667.

[15] E. Hille, *Ueber die Nullstellen der Hermiteschen Polynome*, Jahresber. Deutsch. Math.-Ver., vol. 44(1934), pp. 162–165.

ON BERNOULLI'S NUMBERS AND FERMAT'S LAST THEOREM (SECOND PAPER)

By H. S. Vandiver

1. **Further examination of Fermat's Last Theorem for special exponents.** In the first paper under the present title[1] the writer gave some of the details of the computations which resulted in the proof of Fermat's Last Theorem for all prime exponents l such that $307 < l < 617$, with the exception of 587. At the end of the paper it is stated that the work has been carried out for 587 and since the criteria are found to hold, the theorem is proved for that exponent. The details are as follows. As noted in B.F. (p. 576) the numbers in the set

$$B_1, B_2, \cdots, B_{\frac{1}{2}(l-3)} \tag{1}$$

which are divisible by l when $l = 587$ are B_{45} and B_{46}, so that 587 is irregular and, as in the treatment of irregular primes in B.F., we employ Theorem 1 of that paper which we repeat here for easy reference:

Theorem 1. *Under the assumptions: none of the units E_a $(a = a_1, a_2, \cdots, a_s)$ is congruent to the l-th power of an integer in $k(\zeta)$ modulo $\mathfrak{p}$, where $\mathfrak{p}$ is a prime ideal divisor of p; p is a prime $< (l^2 - l)$ of the form $1 + lk$; and $a_1, a_2, \cdots, a_s$ are the subscripts of the B's in the set (1) which are divisible by l; the relation*

$$x^l + y^l + z^l = 0 \tag{2}$$

is impossible in non-zero integers x, y and z, if l is a given odd prime, and

$$E_n = \prod_{i=0}^{\frac{1}{2}(l-3)} \epsilon(\zeta^{r^i})^{r^{-2in}},$$

$$\epsilon = \left(\frac{(1-\zeta^r)(1-\zeta^{-r})}{(1-\zeta)(1-\zeta^{-1})}\right)^{\frac{1}{2}},$$

r being a primitive root of l and $\zeta = e^{2i\pi/l}$.

Applying this to the case $l = 587$, we find for $r = -10$, $d = 2^{14}$, $p = 8219$, $\rho = 2$ and $n = 45$, ind $E_n(d) \equiv 576 \pmod{587}$ and for $n = 46$, ind $E_n(d) \equiv 60 \pmod{587}$. Here, as in B.F., d is an integer such that $d^l \equiv 1 \pmod{p}$ and ρ is a primitive root of p. Since ind $E_n(d) \not\equiv 0 \pmod{l}$ in the above, the criteria of the theorem are satisfied and Fermat's Last Theorem is proved for $l = 587$.

As noted in B.F. (p. 576) the prime 617 is irregular and B_{10}, B_{87} and B_{169} constitute all the B's in the set (1) which are divisible by l. Then applying Theorem 1, we find for $r = 410$, $d = 3^8$, $p = 4937$, $\rho = 3$, ind $E_{10}(d) \equiv 55$;

Received January 30, 1939.

[1] This Journal, vol. 3(1937), pp. 569–584. This paper will be referred to here as B.F.

ind $E_{87}(d) \equiv 376$; ind $E_{169}(d) \equiv 25 \pmod{617}$. In view of this, Fermat's Last Theorem is proved for all exponents < 619 and the second factor of the class number of $k(\zeta)$ is prime to l for all l's < 619. (Cf. B.F., Theorem 3, p. 581.)

The above-mentioned extensive computations concerning $l = 587$ and $l = 617$ were carried out by M. E. Tittle, who, with M. M. Abernathy and D. H. Lehmer, also directed or carried out all the computations described in B.F.

The prime $l = 617$ just disposed of is a particularly interesting one since it is the first prime so far encountered in our work in which three distinct B's in (1) are divisible by l. Since the criteria hold for this case, there is no indication yet that they will fail because of the number of B's in (1) which are divisible by l.

In B.F. (top of p. 570) we noted that we have persisted in the examination of special exponents in (1) in the hope that, if the criteria of Theorem 1 fail for a particular l, we shall find such an l within the range of our computations. We shall point out here, however, that the possible fact that they might fail for a particular l will not immediately appear when we apply Theorem 1. For, suppose we find that $E_n(d) \equiv 0 \pmod{l}$ for some particular d, l and p; then of course it does not necessarily follow that there will not be another p, say p_1, in the range mentioned in Theorem 1 such that if $d_1^l \equiv 1 \pmod{p_1}$,

$$\text{ind } E_n(d_1) \not\equiv 0 \pmod{l}.$$

But since ind $E_n(d) \not\equiv 0 \pmod{l}$ for every value of l so far tested with the least possible value of p which satisfies $p = 1 + cl$, we should be a little suspicious about any exponent l for which it was found that $E_n(d) \equiv 0 \pmod{l}$ for the first two or three possible values for p, and then it would be in order to subject the cyclotomic field $k(\zeta)$ corresponding to this particular value of l to a special examination. For example, we would try to find out if either of the congruences ($\mu = \frac{1}{2}(l - 1)$)

$$B_{nl} \equiv 0 \pmod{l^3}, \qquad B_{nl-\mu} \equiv 0 \pmod{l^2}$$

held. (Cf. B.F., p. 582.)

2. **Possible extensions of Theorem** 1. This theorem contains the limitation "p is a prime $< (l^2 - l)$ of the form $(1 + kl)$". In the first place it is not known that a prime p of this type exists for every l. Hence we examine the possibility of extending the theorem by widening the range for p. As in a previous paper[2] we consider

$$\omega^l + \theta^l = -\gamma^l, \tag{3}$$

where $\theta \equiv 0 \pmod{\lambda^s}$, $s > 0$, $\lambda = (1 - \zeta)$, and obtain therefrom, if $p = 1 + cl$,

$$(\omega + \theta\zeta^a)^c \equiv (\omega + \theta\zeta^{-a})^c \pmod{\mathfrak{p}},$$

[2] Transactions of the American Mathematical Society, vol. 31(1929), pp. 631–632, relations (22) and (24a).

where $\mathfrak{p}$ is defined as in Theorem 1, with $c = k$. From this we obtain, by expansion, allowing a to range over the set $0, 1, \cdots, l-1$ and eliminating ζ,

$$\binom{c}{s}\omega^{s}\theta^{c-s} + \binom{c}{s+l}\omega^{s+l}\theta^{c-(s+l)} + \cdots \equiv \binom{c}{l-s}\omega^{l-s}\theta^{c-(l-s)} + \binom{c}{2l-s}\omega^{2l-s}\theta^{c-(2l-s)} + \cdots \tag{4}$$

modulo $\mathfrak{p}$, where $s = 0, 1, \cdots, l-1$ and $\binom{c}{h} = 0$ for $h > c$. Taking $c < l - 1$ and $s = c - 1$, assuming that ω is prime to $\mathfrak{p}$, we have $\theta \equiv 0 \pmod{p}$, and this is analogous to the relation (25a) of the paper just mentioned. For larger values of c these congruences (4) are more complicated, but obviously we may state criteria concerning (2) in Case II, based on them. These congruences are related to some congruences given in a previous paper.[3]

Another point of view is to take (2) and obtain therefrom

$$x + \zeta^{a}y = \eta_a \sigma_a^{l},$$

where $y \equiv 0 \pmod{l}$. Now if just one of the B's in the set (1) is divisible by l, say B_n, then it is possible to show that since η_a is primary,

$$\eta_a = E_n^{t}\delta^{l},$$

and since $E^{s-r^{2n}} = \beta^{l}$ (using the Kronecker-Hilbert notation of symbolic powers) we obtain

$$x + \zeta^{ar}y = (x + \zeta^{a}y)^{r^{2n}}\sigma_1^{l},$$

with σ_1 a number in $k(\zeta)$, and therefore

$$(x + \zeta^{ar}y)^{c} \equiv (x + \zeta^{a}y)^{cr^{2n}} \pmod{\mathfrak{p}}.$$

By expansion and eliminating ζ by using various values of a, we obtain another set of congruences involving x and y. By taking r^2, r^3, etc. in lieu of r we obtain numerous relations of this type. If more than one B in (1) is divisible by l, then we may extend this idea, and by taking a certain symbolic power of $(x + \zeta^{a}y)$, we may obtain congruences which become quite complicated when a number of B's are divisible by l. This line of attack seems to have the peculiarity, however, that it yields results depending on the fact that x and y are rational.

3. **Case I of Fermat's Last Theorem and the second factor of the cyclotomic class number.** Take the relation (2) and assume that $xyz \not\equiv 0 \pmod{l}$ and $(x, y, z) = 1$; then

$$x + \zeta^{a}y = \mathfrak{a}^{l}, \tag{5}$$

[3] Proc. Nat'l. Acad. Sci., vol. 15(1929), theorem on p. 45.

where $\mathfrak{a}$ is an ideal in $k(\zeta)$. If we use some results of Pollaczek,[4] it follows that we have

$$x + \zeta y = \eta\omega_1^{c_1}\omega_2^{c_2} \cdots \omega_s^{c_s}\theta^l, \tag{5a}$$

where the ω's are singular numbers; that is, we have $(\omega_i) = \mathfrak{b}_i^l$, where θ is not a principal ideal in $k(\zeta)$; η is a unit and θ an integer in $k(\zeta)$.

Each ω has the property that

$$\omega(\zeta^r) = (\omega(\zeta))^{r^i}\tau^l, \tag{6}$$

r being a primitive root of l and τ a number in $k(\zeta)$. Hence[5] either $\omega\omega_{-1} = \gamma^l$ or $\omega/\omega_{-1} = \xi^l$ according as i is odd or even, ω_h denoting the number obtained from ω by the substitution (ζ/ζ^h). If there is an $\omega = \mathfrak{b}^{ld}$ which satisfies (6) for i odd and which also divides $(x + \zeta y)$, then we proceed in the following way. We introduce l-th power characters defined by

$$\left(\frac{\theta}{\mathfrak{p}}\right) = \zeta^t; \qquad \theta^{(N(\mathfrak{p})-1)/p} \equiv \left(\frac{\theta}{\mathfrak{p}}\right) \pmod{\mathfrak{p}},$$

$\mathfrak{p}$ being an ideal prime in $k(\zeta)$ prime to $l\theta$ and $N(\mathfrak{p}) = p^f$ the norm of $\mathfrak{p}$. Further for $\mathfrak{a}$ an ideal prime to l and θ, let

$$\mathfrak{a} = \mathfrak{p}_1\mathfrak{p}_2 \cdots \mathfrak{p}_s,$$

where the $\mathfrak{p}$'s are ideal primes in $k(\zeta)$; then

$$\left(\frac{\theta}{\mathfrak{a}}\right) = \left(\frac{\theta}{\mathfrak{p}_1}\right)\left(\frac{\theta}{\mathfrak{p}_2}\right) \cdots \left(\frac{\theta}{\mathfrak{p}_s}\right).$$

From (6) $\mathfrak{b}\mathfrak{b}_{-1}$, since i is odd, is a principal ideal, say (α), and we may apply Furtwangler's law of reciprocity and obtain, if $a \not\equiv \pm 1 \pmod{l}$, $a \not\equiv 0 \pmod{l}$,

$$\left(\frac{x + \zeta^a y}{\alpha}\right) = \left(\frac{\alpha, x + \zeta^a y}{\mathfrak{l}}\right)\left(\frac{\alpha}{x + \zeta^a y}\right).$$

From (5)

$$\left(\frac{\alpha}{x + \zeta^a y}\right) = 1,$$

so that

$$\left(\frac{x + \zeta^a y}{\alpha}\right) = \left(\frac{\alpha, x + \zeta^a y}{\mathfrak{l}}\right), \tag{7}$$

$a \not\equiv \pm 1 \pmod{l}$. If we set

$$\left(\frac{\alpha, x + \zeta^a y}{\mathfrak{l}}\right) = \zeta^L, \tag{7a}$$

[4] Math. Zeitschrift, vol. 21(1924), pp. 1–39.

[5] Pollaczek, loc. cit., p. 22.

then, if $N(\beta)$ is the norm of β in $k(\zeta)$,

$$L \equiv -l_1(\alpha)\frac{N(x+\zeta y)-1}{l} + al_1(x+\zeta y)\frac{N(\alpha)-1}{l} + \sum_{s=2}^{l-2}(-1)^{s-1}a^{l-s}l^{(s)}(\alpha)l^{(l-s)}(x+\zeta y). \tag{7b}$$

Also, if

$$\theta = a_0 + a_1\zeta + \cdots + a_{l-2}\zeta^{l-2},$$

then

$$\theta(e^v) = a_0 + a_1 e^v + \cdots + e^{v(l-2)}$$

and

$$l^{(s)}(\theta) = \left[\frac{d^s \log \theta(e^v)}{dv^s}\right]_{v=0}.$$

Now by a result in a previous paper of the writer's[6] we have

$$\frac{N(x+\zeta y)-1}{l} \equiv 0 \pmod{l},$$

and by Furtwangler's result that if c divides z in (2), then

$$c^{l-1} \equiv 1 \pmod{l^2},$$

we obtain, since $\mathfrak{b}$ and $\mathfrak{b}_{-1}$ divide z by hypothesis,

$$\frac{N(\alpha)-1}{l} \equiv 0 \pmod{l}.$$

Now

$$\left(\frac{x+\zeta^a y}{\alpha}\right) = \left(\frac{x+\zeta^a y}{\mathfrak{b}\mathfrak{b}_{-1}}\right) = \left(\frac{(x+\zeta^a y)(x+\zeta^{-a}y)^{l-1}}{\mathfrak{b}}\right) \tag{8}$$

and

$$\left(\frac{x+\zeta^a y}{\mathfrak{b}}\right) = \left(\frac{x+\zeta y+(\zeta^a-\zeta)y}{\mathfrak{b}}\right) = \left(\frac{(\zeta-\zeta^a)y}{\mathfrak{b}}\right). \tag{8a}$$

Similarly, we have

$$\left(\frac{x+\zeta^{-a} y}{\mathfrak{b}}\right) = \left(\frac{(\zeta-\zeta^{-a})y}{\mathfrak{b}}\right). \tag{8b}$$

We also employ the relation[7]

$$\operatorname{ind}\left(\frac{\zeta^t-1}{\zeta-1}\right) \equiv \frac{1}{2}(t-1)\operatorname{ind}\zeta - 2\sum_{n=1}^{l_1}\frac{(t^{2n}-1)\operatorname{ind}E_n(\zeta)}{r^{2n}-1}$$

[6] Proc. Nat'l. Acad. Sci., vol. 15(1929), p. 44.
[7] Kummer, Journal für Math., vol. 56(1859), p. 277.

modulo l, where $l_1 = \frac{1}{2}(l - 3)$ and $(\theta/\mathfrak{b}) = \zeta^{\text{ind}\,\theta}$. Applying this to (7), using (7a), (7b), (8), (8a), and (8b), we have, noting that ind $(\zeta) \equiv 0$,

$$
\begin{aligned}
&2 \sum_{n=1}^{l_1} \frac{((a-1)^{2n} - 1) \operatorname{ind} E_n(\zeta)}{r^{2n} - 1} - 2 \sum_{n=1}^{l_1} \frac{((a+1)^{2n} - 1) \operatorname{ind} E_n(\zeta)}{r^{2n} - 1} \\
&\qquad \equiv \sum_{s=2}^{l-2} (-1)^{s-1} a^{l-s} l^{(s)}(\alpha) l^{(l-s)}(x + \zeta y)
\end{aligned} \tag{9}
$$

modulo l. From this relation we obtain by expansion a congruence which may be put in the form

$$
-4(a^{2n-1} A_{2n-1} + a^{2n-3} A_{2n-3} + \cdots + aA_1) \equiv \sum_{s=2}^{l-2} (-1)^{s-1} a^{l-s} l^{(s)}(\alpha) l^{(l-s)}(x + \zeta y),
$$

modulo l. After dividing by a we may write

$$
\begin{aligned}
&-4A_0 - \sum_{n=2}^{\frac{1}{2}(l-3)} a^{2n-2}(4A_{2n-1} + (-1)^{2n-1} l^{l-2n+1}(\alpha) l^{(2n-1)}(x + \zeta y)) \\
&\qquad \equiv \sum_{s_1=1}^{\frac{1}{2}(l-3)} a^{l-2s_1-2} l^{(2s_1+1)}(\alpha) l^{l-2s_1-1}(x + \zeta y) - a^{l-3} l^2(\alpha) l^{(l-2)}(x + \zeta y).
\end{aligned} \tag{10}
$$

By the Kummer criteria for Fermat's Last Theorem, we have

$$
B_1 l^{(l-2)}(x + \zeta y) \equiv 0 \pmod{l},
$$

and since $B_1 \not\equiv 0 \pmod{l}$,

$$
l^{(l-2)}(x + \zeta y) \equiv 0 \pmod{l}.
$$

Employing this in (10) and setting $a = 2, 3, \cdots, (l - 2)$ in turn, we have $(l - 3)$ congruences, and since the determinant of the powers of the a's is an alternant which is prime to l, since each $a < l$, we have

$$
\begin{aligned}
&A_0 = 0, \qquad 4A_{2n-1} - l^{(l-2n+1)}(\alpha) l^{(2n-1)}(x + \zeta y) \equiv 0 \pmod{l}, \\
&\qquad\qquad\qquad\qquad\qquad\qquad (n = 2, 3, \cdots, \tfrac{1}{2}(l - 3)); \\
&l^{(2s_1+1)}(\alpha) l^{(l-2s_1-1)}(x + \zeta y) \equiv 0 \pmod{l}, \qquad (s = 1, 2, \cdots, \tfrac{1}{2}(l - 3)).
\end{aligned} \tag{11}
$$

The last relations are trivial since

$$
(\alpha) = \mathfrak{b}\mathfrak{b}_{-1},
$$

and hence α belongs to the real field $k(\zeta + \zeta^{-1})$ and

$$
\left[\frac{d^{2t+1} \log \alpha(e^v)}{dv^{2t+1}}\right]_{v=0} \equiv 0 \pmod{l}.
$$

From the relations (11) we obtain for $n = \frac{1}{2}(l - 3)$, by using the actual values of the A's from (9),

(12) $$A_{l-4} = \frac{l-3}{r^{l-3}-1}\operatorname{ind} E_{l_1}(\zeta),$$

(13) $$A_{l-6} = \binom{l-3}{3}\frac{\operatorname{ind} E_{l_1}(\zeta)}{r^{l-3}-1} + \frac{(l-5)\operatorname{ind} E_{l_1-1}(\zeta)}{r^{l-5}-1}.$$

Also, in the Kummer criteria

$$B_n l^{(l-2n)}(x+\zeta y) \equiv 0 \pmod{l}$$

for the solution of (2), let $n = 2, 3, 4, 5, 6$ and 7. These criteria give, as in another paper[8]

$$l^{(l-2n)}(x+\zeta y) \equiv 0 \pmod{l} \qquad (n = 2, 3, 4, 5, 6, 7),$$

and these congruences applied to (11), (12), and (13) give

$$\operatorname{ind} E_{l_1}(\zeta) \equiv \operatorname{ind} E_{l_1-1}(\zeta) \equiv 0 \pmod{l},$$

and similarly, by taking the values of A_{l-8}, etc. in (11), we find

$$\operatorname{ind} E_{l_1-j}(\zeta) \equiv 0 \pmod{l} \qquad (j = 2, 3, 4, 5).$$

We may now follow the argument as given on pages 122 and 123 of the paper last mentioned and obtain the contradiction mentioned there just at the end of paragraph (1), page 123, and *we have another proof of Theorem* 1, *page* 118:– *If* (2) *is impossible in Case* I, *then the second factor of the class number of the cyclotomic field* $k(\zeta)$ *is prime to* l. *The relation* (11) *of the present paper is different, however, from any of those given in the former paper, and in the main the ideas given here constitute extensions of those employed there.* Now, instead of taking

$$\left(\frac{x+\zeta^a y}{\alpha}\right)$$

and treating it as above, we might have used

$$\left(\frac{x+\zeta^a y}{\beta}\right),$$

where $(\beta) = \mathfrak{b}^{s-r^{2i+1}}$. We would thereby obtain

(14) $$\left(\frac{x+\zeta^a y}{\beta}\right) = \left(\frac{(x+\zeta^{r_1 a} y)^r (x+\zeta^a y)^{l-r^{2i+1}}}{\mathfrak{b}}\right),$$

where $rr_1 \equiv 1 \pmod{l}$.

We may then treat (14) as (8) was treated and obtain a set of congruences analogous to (9), but more general.

By employing (5a) and (6) and proceeding as in another paper of this writer's,[9]

[8] Vandiver, Bull. Amer. Math. Soc., vol. 40(1934), p. 122.

[9] *On criteria for singular integers in a cyclotomic field*, Proc. Nat'l. Acad. Sci., vol. 24 (1938), pp. 330–333.

we obtain

$$\prod_{d=1}^{l-1} (x + \zeta^d y)^{d^{l-2s-2}} = \omega_k^{c_k} \gamma^l, \tag{15}$$

where

$$\omega_k(\zeta^r) = (\omega_k(\zeta))^{r^{2s+1}} \sigma^l,$$

σ being a number in $k(\zeta)$, with s in the set $1, 2, \cdots, \frac{1}{2}(l-3)$.

4. **Criteria involving Bernoulli numbers.** In B.F., pp. 582–583, we mentioned a Theorem 4, giving criteria for Fermat's Last Theorem involving the assumption that none of the Bernoulli numbers

$$B_{nl} \qquad (n = 1, 2, \cdots, \tfrac{1}{2}(l-3))$$

is divisible by l^3. Elsewhere[10] we have stated that if (2) holds in Case I, then

$$B_s \equiv 0 \pmod{l^2} \qquad (s = n_i\mu - i;\ i = 2, 3, 4, 5, 6;\ \mu = \tfrac{1}{2}(l-1)),$$

where each n ranges over all positive integers. In the reference to B.F. just mentioned (middle of page 582) it was shown that if two of the numbers in the set

$$B_n, B_{n+\mu}, \cdots, B_{n+(l-1)\mu} \tag{16}$$

are divisible by l^2, then all of them are. Also, a necessary condition that $k(\zeta)$ contain an ideal belonging to the exponent l^2 is that one of the numbers $B_{\frac{1}{2}(tl+1)}$ $(t = 1, 3, \cdots, l-4)$ be divisible by l^2.

We shall now consider methods for testing these various criteria in special cases. In a previous paper[11] the writer described a method for obtaining the least residue of B_{nl} modulo l^3 where $B_n \equiv 0 \pmod{l}$, using the formula

$$\frac{(-1)^{n-1} B_{nl}(2^{2nl} - 1)}{2nl} \equiv 1^{2nl-1} + 3^{2nl-1} + \cdots + (l-2)^{2nl-1},$$

modulo l^2. Beeger[12] obtained the congruence

$$(-1)^{i-1} \frac{B_{nl}}{2nl} \equiv \sum_{s=1}^{i} (-1)^{(s-1)(\mu+1)} \binom{2n}{s-1} \binom{2n-s}{i-s} \frac{B_{n+(s-1)\mu}}{2n + 2\mu(s-1)}$$

modulo l^i, $2n \not\equiv 0 \pmod{\overline{l-1}}$.

This gives for $i = 2$

$$B_{nl} \equiv -\frac{l}{2n-1} ((2n-1)^2 B_n - (-1)^{\mu}(2n)^2 B_{n+\mu})$$

[10] Bull. Amer. Math. Soc., vol. 40(1934), p. 124.

[11] Trans. Amer. Math. Soc., vol. 31(1929), p. 639.

[12] *On some new congruences in the theory of Bernoulli's numbers*, Bull. Amer. Math. Soc., vol. 44(1938), p. 688.

modulo l^3. This was employed by Beeger to show that $B_{nl} \not\equiv 0 \pmod{l^3}$ for any B_n such that $B_n \equiv 0 \pmod{l}$ and for l any prime < 211 (with several exceptions which were not tested). Emma Lehmer[13] derived the congruence

$$\text{(19)} \qquad \frac{B_n(6^{2n-1} + 3^{2n-1} + 2^{2n-1} - 1)}{2n} \equiv \sum_{r=1}^{[p/6]} (l - 6r)^{2n-1} \pmod{l^2},$$

where $l > 5$ and $[x]$ is the greatest integer in x.

Putting these methods and results together, we evolve the following convenient scheme for finding which numbers in the set (16) are divisible by l^2. Take a congruence employed by Pollazcek[14]

$$\text{(20)} \qquad B'_{n+y\mu} \equiv yB'_{n+\mu} - (y - 1)B'_n \pmod{l^2},$$

where

$$B'_i = \frac{(-1)^i B_i}{i}$$

and $y \geqq 0$. Suppose first, using (19), we find that

$$\text{(21)} \qquad B'_{n+\mu} \equiv B'_n \pmod{l^2};$$

then (20) reduces to

$$B'_{n+y\mu} \equiv B'_n \pmod{l^2},$$

so that there is no member of the set (16) divisible by l^2 unless all are.

Suppose next that (21) does not hold; then determine $0 \leqq y < l$ from the linear congruence

$$y\,\frac{B'_{n+\mu} - B'_n}{l} + \frac{B'_n}{l} \equiv 0 \pmod{l},$$

whence, from (20), $B'_{n+y\mu} \equiv 0 \pmod{l^2}$, and this is the unique B in (16) which is divisible by l^2. For, otherwise all the B's in (16) are divisible by l^2, and (21) holds, contrary to hypothesis.

As we have already noted, the conditions $B_{nl-\mu} \equiv 0 \pmod{l^2}$ and $B_{nl} \equiv 0 \pmod{l^3}$ occur in connection with various questions concerning $k(\zeta)$. The relation (21) also appears in theorems concerning the field defined by a primitive l^2-th root of unity.[15] In employing (19) it may be most convenient when l is

[13] Annals of Math., vol. 39(1938), p. 352, relation (13).

[14] Math. Zeitschrift, vol. 21(1924), p. 36. This congruence was generalized by Beeger, loc. cit., p. 684.

[15] Cf. Pollaczek, loc. cit., p. 29; Morishima, Japanese Journal of Math., vol. 11(1935), p. 239.

large to write it in the form

$$(22)\quad \frac{B_n(6^{2n-1} + 3^{2n-1} + 2^{2n-1} - 1)}{2n} \equiv \sum_{r=1}^{[l/6]} (6^{2n-1}r^{2n-1} + (2n-1)6^{2n-2}r^{2n-2}l)$$

modulo l^2; for, as indicated elsewhere,[16] it is convenient in finding the least residue of $(ak)^{n-1}$ modulo l^2 to obtain it from the least residues of a^{n-1} and k^{n-1}. The least residue of the coefficient of l in the right member of (22) may be determined easily if we use Jacobi's table of indices.

UNIVERSITY OF TEXAS.

[16] Vandiver, Trans. Amer. Math. Soc., vol. 31(1929), p. 641.

NOTE ON TOPOLOGICAL MAPPINGS

By J. H. Roberts

E. W. Miller[1] has given an example of an acyclic curve M such that if f is any topological mapping of M into a subset of itself, then $f(M) = M$. R. Baer[2] has given an example of an acyclic curve M such that if f is a topological function and $f(M) = M$, then f is the identity. Neither of these examples has *both* of the properties mentioned above. O. Hamilton[3] has raised the question as to whether or not any acyclic curve has both the above properties. The present paper answers this question in the affirmative by describing a compact acyclic continuous curve H such that the only topological function mapping H into a subset of itself is the identity.

Now Menger's[4] "universal tree of order 4" is made up as follows: (1) There is a single interval S which is called the interval of the "0-th degree". (2) For each point P of a countable set T_0 dense on S, but not containing an end-point of S, there are two intervals having P as end-point, these intervals being of the 1-st degree. (3) In general, for every $n \geqq 0$ there is a countable set T_n dense on every interval of the n-th degree and for each point P of T_n there are two intervals having P as end-point, these intervals being the intervals of the $(n + 1)$-th degree. (4) The curve M is the sum of all the intervals of all the different degrees, plus all limit points of this sum.

Our curve H will be defined as a subset of Menger's curve M. To get H we modify M in this way: Having decided that a certain interval I of degree r (in M) is to be in H, we may wish to have only *one* interval of degree $r + 1$ for each of the junction points on I. In this case we select arbitrarily (to be a part of H) one of the two intervals of degree $r + 1$ ending in each junction point on I. In the future we will indicate this by writing "the junction points on I are to be of order 3 in H".

It is convenient to use the following notation: Suppose P is a junction point of M on an arc of degree r. Then an arc I of degree $> r$ is said to "join on through P" if P separates I (or $I - P$) from S (or $S - P$).

We now set up a 1-1 correspondence between the set of all finite permutations of positive integers and the integers of the form 2^k. Let $x_{ij\ldots k}$ be the

Received March 9, 1939.

[1] *The Zarankiewicz problem*, Bull. Amer. Math. Soc., vol. 38(1932), pp. 831–834.

[2] *Beziehungen zwischen den Grundbegriffen der Topologie*, Sitzungsberichte der Heidelberger Akademie der Wissenschaften, 1929, no. 15.

[3] *Fixed points under transformations of continua*, Trans. Amer. Math. Soc., vol. 44(1938), pp. 18–24; especially p. 24.

[4] *Kurventheorie*, Leipzig, 1932, p. 318.

integer associated with the permutation $ij \cdots k$. The peculiar property of the integers of the form 2^k of which we make use here is the following:

I. *No one of them is equal to a sum of any of the remaining ones.*

Let $P_1, P_2, \cdots$ be the junction points of M on S. This interval is to be in H, and the points $P_1, P_2, \cdots$ are all to be of order 3 in H. Let I be an interval of degree r which is in H and which joins on through P_i. If $r < x_i - 1$, then the junction points on I are all to be of order 3 in H. If $r = x_i - 1$, these points are all to be of order 4. Let $\{P_{ij}\}$ $(j = 1, 2, \cdots)$ be a countable set dense on every arc of degree $x_i - 1$ which is in H and which joins on through P_i. These points are of order 3 in H. Suppose now the arc I of degree r $(\geqq x_i)$ is in H and joins on through P_{ij}. Then if $r < x_i + x_{ij} - 1$, the junction points on I are to be of order 3 in H. If $r = x_i + x_{ij} - 1$, these junction points are to be of order 4. We can continue in this way to specify the order of the junction points (to be 3 or 4) so that the following properties obtain:

(1) On any given interval all junction points (except possibly one end-point of this interval) are of the same order.

(2) Suppose $ij \cdots kl$ is a permutation of positive integers, I is an interval of degree r which is in H and which joins on through the point $P_{ij\cdots kl}$.

Let

$$q_1 = x_i + x_{ij} + \cdots + x_{ij\cdots k}, \qquad q_2 = q_1 + x_{ij\cdots kl}.$$

If $q_1 \leqq r < q_2 - 1$, then the junction points on I are of order 3 in H.

If $r = q_2 - 1$, these junction points are of order 4. The point set

$$\{P_{ij\cdots klm}\} \qquad (m = 1, 2, 3, \cdots)$$

is dense on every interval of degree $q_2 - 1$ in H which joins on through $P_{ij\cdots kl}$.

This completes the definition of H.

We note for future use the following property:

II. *Suppose $Z_1, Z_2, \cdots$ is a sequence of junction points, such that* (1) *the arc Z_iZ_{i+1} (in H) is a subset of a single interval, and* (2) *if the degree of the interval containing Z_1Z_2 is $s + 1$, then the degree of the interval containing Z_iZ_{i+1} is $s + i$. Then if Z_h and Z_k are of order* 4, *and $k > h$, $k - h$ is a sum of numbers of the set $\{x_{ij\cdots l}\}$.*

Suppose now that f is a topological function mapping H into a subset of H. We wish to show that for every $P \epsilon H$ we have $f(P) = P$. Since the set $\{P_{ij\cdots l}\}$ is dense in H, it will be sufficient to show that for any $ij \cdots k$ we have $f(P_{ij\cdots k}) = P_{ij\cdots k}$. Suppose that this is false and that $f(P) = Q \neq P$, where P is some $P_{ij\cdots k}$.

Select mutually exclusive open sets U and V in H containing P and Q respectively, and such that $f(U) \subset V$. Let PX be an arc which is a subset of an interval in H and let r be the degree of this interval. Then there exists[5] a point

[5] This follows easily from Menger's "n-Beinsatz", ibid., p. 214. The immediate assumption is the following theorem: If H is an acyclic continuous curve, B is a point of H of order n, and A_iB is an arc in H, for $i = 1, 2, \cdots, n, n + 1$, then for some i and j $(i \neq j)$ there is a point X such that the arc XB in H is a subarc both of A_iB and of A_jB.

Y_1 on the arc PX such that (1) Y_1 is a junction point of M, (2) PY_1 is in U, and (3) if $f(PY_1) = QZ_1$, then QZ_1 is a subset of a unique interval of H, say one of degree s.

Then on the interval (or on one of the two intervals) in H of degree $r + 1$ having Y_1 as an end-point there exists a point Y_2 such that (1) Y_2 is a junction point of M, (2) Y_1Y_2 is in U, and (3) if $f(Y_1Y_2) = Z_1Z_2$, then Z_1Z_2 is a subset of an interval of M of degree $s + 1$.

This process can be continued indefinitely. Thus there exists an infinite sequence of arcs $PY_1, Y_1Y_2, Y_2Y_3, \cdots$ such that for each i, we have

(1) Y_i is a junction point of M;

(2) Y_iY_{i+1} is in U and is a subset of an interval of degree $r + i$;

(3) $f(Y_i) = Z_i$, and the arc Z_iZ_{i+1} is a subset of an interval of degree $s + i$.

But the following consideration leads us to a contradiction: Among the points $Y_1, Y_2, \cdots$ let Y_h and Y_k be respectively the first and second of order 4. Then $k - h$ is some integer $x_{ij\cdots l}$ associated with a point $P_{ij\cdots l}$ in U. Now clearly Z_h and Z_k, being images of Y_h and Y_k, respectively, are of order 4. But then by Property II $k - h$ is a sum of numbers of the form $x_{ij\cdots l}$ associated with points in V. Since $U \cdot V = 0$, $P_{ij\cdots l}$ is not in V.

Thus by Property I we have reached a contradiction.

DUKE UNIVERSITY.

CONTRIBUTIONS TO THE THEORY OF GROUPS OF FINITE ORDER

By Oystein Ore

The present paper contains a number of results of diverse nature in the theory of finite groups. One may say that the guiding principle is the application of structure theory to the theory of groups. In two recent papers[1] I have already shown that this method is very useful for various investigations in group theory. In the present paper this point of view is particularly important in the study of *non-normal* chains of subgroups of groups, a field which seems to be untouched until now. One of the main problems which is solved by this structural approach is the determination of all groups in which *every* chain of subgroups, with each term maximal, but usually not normal, in the preceding shall have the Jordan-Hölder property that the indices are the same in some order.

Let us also make the following general remark. Groups are ordinarily defined by their elements, or, equivalently in structural terms, each subgroup is the union of cyclic groups, and hence the group properties are naturally stated by means of element properties. By dualizing this process one is led to the investigation of the properties of a group in relation to its maximal subgroups and several of the results of this paper may be said to belong to this category.

In the first chapter various properties of permutable groups are derived and particularly the existence of permutable decompositions is investigated. It is shown that if a group has the property that all maximal subgroups are permutable, they are all normal and the group is nilpotent. The concept of quasi-normality introduced in the preceding papers is studied further, and it is shown that in several cases normality and quasi-normality are identical concepts. In particular, no simple groups can contain quasi-normal subgroups.

In the second chapter properties of normal decompositions as union and the dual decomposition as cross-cut are considered and the relation to the representations of the group as a permutation group is pointed out. Here one also finds the determination of the ϕ-group of a group.

In the third chapter properties of arbitrary non-normal chains of subgroups are deduced. It is shown that to any complete chain there exists a chain with the same index type passing through any prescribed principal chain. This result reduces the study of various properties of such chains to the case of simple groups.

Received March 24, 1939.

[1] Oystein Ore, *Structures and group theory*, I and II, this Journal, vol. 3(1937), pp. 149–174; vol. 4(1938), pp. 247–269. These papers will be quoted in the following as Ore I, II.

When this theory is applied to solvable groups in Chapter IV, it follows immediately that all consecutive indices in such groups are powers of primes and each subgroup is permutably contained in any immediately preceding group. The properties of maximal groups in solvable groups are particularly interesting. Each such group belongs to a unique normal subgroup which determines it except for isomorphism. Any two maximal subgroups are conjugate or permutable, and if p divides the group order, there exists a maximal group whose index is a power of p. This last statement is equivalent to a result of P. Hall.

In the last chapter one finds the solution of the problem of finding all groups in which any maximal chains have the property that the indices of two chains are the same in some order. It is shown that these groups must be solvable. They form a class of groups which may also be characterized by various other properties, for example, the properties that all consecutive indices be primes, that there exist a complete principal chain, or that the subgroups and quotient groups have subgroups of every possible order.

Chapter I. Permutability

1. **Permutable groups.** In the following we shall consider the subgroups of a group G. When A and B are two such subgroups, we shall denote their *union* by $A \cup B$ and their *cross-cut* by $A \cap B$.

The two subgroups shall be said to be permutable if to every a_1 in A and b_1 in B there can be found a second pair a_2 and b_2 such that $a_1 \cdot b_1 = b_2 \cdot a_2$.

Let us recall the following properties of permutable subgroups:[2]

If A is permutable with B and C, then A is permutable with $A \cup C$.

A group is permutable with all its subgroups.

If $\bar{B} \supset B$ and A is permutable with B, then B is permutable with $A \cap \bar{B}$.

If $\bar{A} \supseteqq A$ and $\bar{B} \supseteqq B$ and A and B are permutable, then $\bar{A} \cap B$ and $A \cap \bar{B}$ are permutable.

One of the fundamental properties of the permutable groups is the

Dedekind Relation. *Let A and B be permutable and $C \supset A$. Then* $C \cap (A \cup B) = A \cup (C \cap B)$.

One can express the condition for two groups to be permutable as follows:

Theorem 1. *Let*

$$A = \sum a_i T, \qquad B = \sum b_j T$$

be the co-set expansions of two groups with respect to a common subgroup T. The necessary and sufficient condition for two groups to be permutable is that there exist relations

$$a_i \cdot b_j = b_k \cdot a_l \cdot t \tag{1}$$

for every pair a_i, b_j.

[2] Ore I, Chapter 2.

Proof. An arbitrary a in A and b in B may be written in the form $a = a_i \cdot t$, $b = b_j \cdot t'$, where t and t' belong to T. When (1) holds, one obtains $a \cdot b = a_i \cdot t \cdot b_j \cdot t' = a_i \cdot b_k \cdot t'' = b_l \cdot a_m \cdot t''' = b' \cdot a'$, and the converse follows similarly.

Now let A and B be permutable and let us write

$$M = A \cup B, \qquad D = A \cap B.$$

If then

$$B = \sum b_i D \tag{2}$$

is a co-set expansion of B with respect to D, then

$$M = \sum b_i A \tag{3}$$

is the co-set expansion of M with respect to A. Even if A and B are not permutable, the right-hand co-sets in (3) are all distinct, and hence one always has the index relation

$$[(A \cup B) : A] \geqq [B : (A \cap B)] \tag{4}$$

and the equality holds only when A and B are permutable.

The correspondence

$$b_i D \rightleftarrows b_i A \tag{5}$$

is a one-to-one correspondence between the elements in the two quotient systems

$$A \cup B/A \rightleftarrows B/A \cap B. \tag{6}$$

From this correspondence one easily derives a set of others:[3]

Let $\bar{A} \supseteqq A$, $\bar{B} \supseteqq B$.

If A is permutable with $\bar{A} \cap \bar{B}$ and $\bar{A} \cap B$ and B is permutable with $\bar{A} \cap \bar{B}$ and $A \cap \bar{B}$, then

$$A \cup (\bar{A} \cap \bar{B})/A \cup (\bar{A} \cap B) \rightleftarrows B \cup (\bar{B} \cap \bar{A})/B \cup (\bar{B} \cap A).$$

If $A \cup B$ is permutable with $\bar{A}$ and $\bar{B}$, then

$$\bar{A} \cap (A \cup \bar{B})/\bar{A} \cap (A \cup B) \rightleftarrows \bar{B} \cap (B \cup \bar{A})/\bar{B} \cap (B \cup A).$$

If A is permutable with $\bar{B}$ and B, and B is permutable with $\bar{A}$ and A, then the last two correspondences are identical and one also has

$$\bar{A} \cap \bar{B}/\bar{A} \cap \bar{B} \cap (A \cup B) \rightleftarrows A \cup B \cup (\bar{A} \cap \bar{B})/A \cup B.$$

The correspondence (5) does not usually give a correspondence between the subgroups of the two quotient structures (6). We shall now analyze this side of the correspondence further. It is seen that if $\bar{A} = \sum \bar{b}_i A$ is a co-set expansion of any subgroup of M containing A, then the multipliers $\bar{b}_i$ generate a group $\bar{B}$ containing D and such that the index relation

$$[\bar{A} : A] = [\bar{B} : D] \tag{7}$$

[3] Ore I, Chapter 2.

holds. This group B may also be given explicitly

$$\bar{B} = B \cap A, \tag{8}$$

and it is seen to be permutable with A.

Conversely, let $\bar{B} = \sum \bar{b}_i D$ be some subgroup of B containing D. Then by (5) $\bar{B}$ corresponds to the co-sets $\sum \bar{b}_i A$. The smallest group containing these co-sets is obviously $\bar{A} = A \cup \bar{B}$, but the index relation (7) will not hold except when $\bar{B}$ is permutable with A. Now to $\bar{A}$ corresponds conversely as in (8) $\bar{\bar{B}} = B \cap (A \cup \bar{B})$, and hence $\bar{\bar{B}}$ is a group in B/D permutable with A and containing $\bar{B}$. It is also the smallest such group because if $B_1 \supset \bar{B}$ is another, then one finds

$$B_1 = B \cap (A \cup B_1) \supseteqq B \cap (A \cup \bar{B}) = \bar{\bar{B}}.$$

In this manner the groups in B/D are distributed into systems, each subgroup corresponding to the least subgroup which contains it and is permutable with A.

Let us finally observe that if B_1 and B_2 are subgroups of B/D permutable with A, then $B_1 \cup B_2$ and $B_1 \cap B_2$ are permutable with it. In the case of the union it is obvious and for the cross-cut it follows from the representation

$$\begin{aligned} B_1 \cap B_2 &= B \cap (A \cup B_1) \cap (A \cup B_2) \\ &= B \cap (A \cup (B_2 \cap (A \cup B_1))) = B \cap (A \cup (B_1 \cap B_2)). \end{aligned}$$

We can summarize these results as follows:

THEOREM 2. *When A and B are permutable groups, then there exists a strong structure isomorphism between the quotient structure $A \cup B/A$ and that substructure of $B/A \cap B$ which consists of those subgroups which are permutable with A.*[4]

2. **Permutable maximal groups.** We shall prove here a few facts about the conjugate groups of permutable groups and show first

THEOREM 3. *Let A and B be permutable. Any conjugate of A in $M = A \cup B$ is then permutable with any conjugate of B in the same group.*

Proof. It is sufficient to prove that any conjugate of A is permutable with B. Now any such conjugate has the form bAb^{-1} and the permutability of this group with B follows by transforming the relation $a_1b_1 = b_2a_2$ by b.

Another property is the following:

THEOREM 4. *Two permutable groups A and B cannot be conjugate in their union.*

[4] A similar investigation has been made for structure of equivalence relations by Paul Dubreil and Mme. Dubreil-Jacotin, *Propriétés algébriques des relations d'équivalence*, Comptes Rendus, vol. 205(1937), pp. 704–706; *Propriétés algébriques des relations d'équivalence; théorèmes de Schreier et de Jordan-Hölder*, Comptes Rendus, vol. 205, pp. 1349–1351.

Proof. Let us suppose that there exists an element $m = a_0b_0$ in M such that $(a_0b_0)B(a_0b_0)^{-1} = A$. Then one finds

$$B = b_0Bb_0^{-1} = a_0^{-1}Aa_0 = A.$$

This theorem has the following interesting application:

THEOREM 5. *If all maximal subgroups of a group G are permutable, then all maximal subgroups are normal.*

We shall show later that this is a characteristic property of the nilpotent groups.

Theorem 4 can be extended as follows:

THEOREM 6. *Let A and B be permutable groups. If a subgroup B_1 of B is conjugate to a subgroup A_1 of A in $A \cup B$, then both are conjugate to a subgroup of $A \cap B$.*

Proof. Since one has $(a_0b_0)B_1(a_0b_0)^{-1} = A_1$, it follows that $b_0B_1b_0^{-1} = a_0^{-1}A_1a_0 \subset A \cap B$.

3. **Existence of permutable decompositions.** We shall say that a group G is *permutably decomposed* if $G = A \cup B$, where A and B are permutable, and in this case we say that A and B are *permutably contained* in G. We shall now study some conditions for one group to be permutably contained in another.[5] From the preceding it is obvious that if A is permutably contained in G, then all its conjugates have the same property.

Now for the moment let A and B be arbitrary (not necessarily permutable) groups, and let i_A, i_B denote the indices of A and B in $M = A \cup B$. Furthermore, let n_M, n_D, n_A, n_B denote the orders of the corresponding groups. From the relations

$$n_M = i_A \cdot \frac{n_A}{n_D} \cdot n_D = i_B \cdot \frac{n_B}{n_D} \cdot n_D$$

follows that n_D divides n_M/i_A and n_M/i_B, and hence it divides $n_M/[i_A, i_B]$, where the bracket denotes the least common multiple.

On the other hand it follows from the relation (4) that

$$i_A \geqq \frac{n_M}{i_B} \cdot n_D^{-1},$$

and hence we can state

THEOREM 7. *Let A and B have the indices i_A and i_B in $A \cup B$ and let n_M and n_D be the orders of $A \cup B$ and $A \cap B$. Then*

[5] Investigations on this problem have been made particularly by E. Maillet: *Sur les groupes échangeables et les groups décomposables*, Bulletin Soc. Math. de France, vol. 28 (1900), pp. 7–16.

$$\frac{n_M}{[i_A, i_B]} \geqq n_D \geqq \frac{n_M}{i_A \cdot i_B} \tag{9}$$

and n_D divides the upper bound.

The lower bound for n_D is attained if and only if A and B are permutable. Hence we can state

THEOREM 8. *Let A and B be subgroups with relatively prime indices in a group G. Then A and B are permutable and $G = A \cup B$.*

From this theorem follows immediately

THEOREM 9. *Let A be a group of prime power index p^n. Then A is permutable with every Sylow group S_p corresponding to p and $G = S_p \cup A$.*

This theorem shows that any subgroup of prime power index p^n is permutably contained in G except possibly when G is a p-group.

Theorem 9 may be extended as follows:

Let G have a subgroup A of index n_1 and let $N = n \cdot m$ be a factorization of the group order into two relatively prime factors, where n_1 divides n. If G has a subgroup B of order n, then A is permutably contained in G and $G = A \cup B$.

Let us finally mention a fact which will be useful later.

THEOREM 10. *Let A be a subgroup of prime index p in G. Then there exists a cyclic group $\{b\} = B$ such that it is permutable with A and $G = A \cup B$, $b^p \subset A$.*

Proof. Let S_p be a Sylow group of G corresponding to p. The cross-cut $S_p \cap A$ must then have the index p in S_p since A and S are permutable. This implies also that $S_p \cap A$ is normal in S_p and any element b in S_p but not in $S_p \cap A$ will satisfy the conditions.

4. **Normal subgroups in permutable groups.** An important problem in connection with permutable decompositions is the question when the existence of a permutable decomposition of a group implies the existence of a normal subgroup.

THEOREM 11. *Let A and B be permutable. Those elements a_0 in A for which ba_0b^{-1} belongs to A for every b in B form a subgroup A_0 which is normal in $A \cup B$.*[6]

Proof. It is obvious that the elements a_0 form a group A_0 and that A_0 is invariant by transformation with elements of B. To show that A_0 is normal in A, let a be any element of A and let us write $\bar{a} = aa_0a^{-1}$. Then for any b

$$b\bar{a}b^{-1} = bab^{-1} \cdot ba_0b^{-1} \cdot (bab^{-1})^{-1} = bab^{-1} \cdot a_0'(bab^{-1})^{-1}.$$

Now let $bab^{-1} = a_2b_2$. Then one finds

$$b\bar{a}b^{-1} = (a_2b_2)a_0'(a_2b_2)^{-1} = a_2 \cdot a_0'' \cdot a_2^{-1} \subset A$$

and $\bar{a}$ belongs to A_0.

The following theorem has very useful applications.

[6] Ore II, Chapter 3.

THEOREM 12. *Let A and B be permutable and let B_0 be a normal subgroup of B which is also contained in $A \cap B$. Then A contains a subgroup $A_0 \supset B_0$ which is normal in $A \cup B$.*

Proof. The conjugates A_i of A in $A \cup B$ are all obtained when A is transformed by the elements of B. Since A contains B_0 and B_0 is normal in B, all conjugates A_i also contain B_0. This shows that the normal subgroup A_0 of $A \cup B$ which is defined by the cross-cut of all A_i must also contain B_0.

An important special case is

THEOREM 13. *Let A and B be permutable and B Abelian. If $A \cap B$ is not the unit group, then A contains a normal subgroup of $A \cup B$.*

5. **Application of a theorem of Frobenius.** We shall now give another case in which the existence of a permutable decomposition implies the existence of a normal subgroup. The method of proof depends on the theorem of Frobenius that if k divides the group order, then the number of solutions of the equation $x^k = e$ is a multiple of k. The method is analogous to a method used by Burnside[7] in the determination of all groups with cyclic Sylow subgroups.

THEOREM 14. *Let A and B be permutable groups such that $A = \{a\}$ is cyclic of prime power order p^α and n_B is relatively prime to $p(p-1)$. Then B is normal in $M = A \cup B$.*

Proof. The number of solutions of the equation

$$x^{p^{\alpha-1}\cdot n_B} = e \tag{10}$$

must be of the form $r \cdot p^{\alpha-1} \cdot n_B$, where $r < p$. The number of elements not satisfying (10) is then $(p-r)p^{\alpha-1}\cdot n_B$, and we shall prove that $r = 1$ by showing that this number must be divisible by $p - 1$.

There are $p^{\alpha-1}(p-1)$ powers of a not satisfying (10). Similarly, each conjugate of a gives rise to $p^{\alpha-1}(p-1)$ different elements of order p^α. Now any element in M can be written uniquely in the form $m = \bar{a}\cdot\bar{b}$, where $\bar{a}$ and $\bar{b}$ are permutable elements (powers of m) and where the order of $\bar{a}$ is a power of p and the order of $\bar{b}$ is a divisor of n_B. In order that such an element shall not satisfy (10) it is necessary and sufficient that the order of $\bar{a}$ be p^α, and hence it is contained as a generating element in a conjugate of $\{a\}$. Now any element $\bar{b}$ permutable with $\bar{a}$ is also permutable with all powers of $\bar{a}$, and hence the elements not satisfying (10) fall into disjoint classes each containing a number of elements divisible by $p - 1$.

By induction one proves that the number of solutions of $x^{p^{\alpha-i}\cdot n_B} = e$ is equal to $p^{\alpha-i}\cdot n_B$. For $i = \alpha$ it follows that $x^{n_B} = e$ has n_B solutions, and this implies obviously that B is normal.

6. **Quasi-normality.** On the basis of permutability conditions various types of normality may be defined. The property of being *permutably contained*

[7] W. Burnside, *Theory of Groups of Finite Order*, 2d ed., pp. 163–164.

may be considered as a weak normality property. The concept of *quasi-normality* is the strongest type of normality short of actual normality. We say that A is quasi-normal in G if A is permutable with every subgroup of G.[8] This means that for any g in G and a in A we have

$$g \cdot a = a' \cdot g^n. \tag{11}$$

If A is quasi-normal in G, then any conjugate of A by any automorphism of G is also quasi-normal.

Let

$$G = \sum g_i A \tag{12}$$

be a co-set expansion of G with respect to A. If A is quasi-normal in G, then all g_i must satisfy permutability relations of the form (11). It does not follow conversely, however, that A is quasi-normal when these hold. The property of having such a permutable representative system g_i must therefore be considered as a weaker normality condition. It follows from Theorem 10 that every group of prime index has a permutable representative system.

By combining Theorems 13 and 10 we obtain

THEOREM 15. *Let A be a group of prime index p in G. Then G is permutably decomposable $G = A \cup B$, where $B = \{b\}$ is cyclic of prime power order p^α. Here either $\alpha = 1$ and $A \cap B = E$ or A contains a normal subgroup of G.*

Let us finally deduce some properties of quasi-normal subgroups. We prove first

THEOREM 16. *Any maximal quasi-normal subgroup is normal.*

Proof. The union of two quasi-normal subgroups is again quasi-normal, and hence the union of a maximal quasi-normal subgroup A and one of its conjugates is equal to G. This is, however, not possible according to Theorem 4, so A must be normal. A consequence of Theorem 16 is

THEOREM 17. *A quasi-normal subgroup of prime index is normal.*

One may define a group G to be *quasi-solvable* when there exists a chain $G \supset A \supset B \supset \cdots \supset E$, where each group is maximal and quasi-normal in the preceding. From Theorem 16 it follows that the chain must be a composition chain and each index a prime.

THEOREM 18. *A quasi-solvable group is solvable.*

From Theorem 16 one also concludes

THEOREM 19. *Any quasi-composition chain in a group is a composition chain and any quasi-normal group occurs in some composition chain.*

[8] Ore I, Chapter 2.

It follows also that a simple group cannot contain any quasi-normal subgroups. A theorem related to the preceding is the following theorem of Frobenius:[9]

THEOREM 20. *Let p be the smallest prime dividing the order of a group G. Then a subgroup of index p is normal.*

Proof. It is sufficient to show that such a subgroup A is quasi-normal. According to Theorem 15 there exists an element b permutable with A and such that b^p is contained in A. Any element b_1 not in A must have the order p with respect to A and one must have relations $b_1^k = a_k \cdot b^l$ $(k = 1, 2, \cdots, p - 1)$, where also l must run through the same numbers in some order. For any element a in A one therefore finds $b_1 \cdot a = a_1 \cdot b^{i_1} \cdot a = a_2 b^{i_2} = a_3 b_1^{i_3}$, and A is quasi-normal.

Chapter II. Normal decompositions

1. **Properties of normal decompositions.** We shall begin by recalling a few of the properties of *normal union decompositions.*[10] Any finite group G has normal decompositions

$$G = A_1 \cup A_2 \cup \cdots \cup A_n, \tag{1}$$

where the A_i are normally indecomposable in G; i.e., there exists no representation $A = B \cup C$, where B and C are normal in G and proper subgroups of A. The condition for A_i to be normally indecomposable in G is that it contain a unique maximal group M_i normal in G. For those A_i which occur in a decomposition (1) one finds that the quotient group

$$P_i = A_i/M \tag{2}$$

is a simple group, and we shall say that A_i *belongs* to this simple group P_i. It is convenient to distinguish between two types of indecomposable groups A_i. We say that A_i is of *Abelian type* if P_i is a cyclic group of prime order and A_i is of *non-Abelian type* when P_i is non-commutative.

One can always suppose that a decomposition (1) is reduced; i.e., it contains no superfluous terms and no A_i can be replaced by a smaller group also normal in G. Then the main theorem on irreducible decompositions (1) states that any other such decomposition $G = B_1 \cup B_2 \cup \cdots \cup B_m$ contains the same number of terms and any one group in one decomposition may be replaced by a suitably chosen group in the other.

Let us introduce the following notations. We write $\bar{A}_i = A_1 \cup \cdots \cup A_{i-1} \cup A_{i+1} \cup \cdots \cup A_n$ and one finds that the groups $C_i = M_i \cup \bar{A}_i$ are maximal normal subgroups of G and $G_i = G/C_i \cong P_i$. If C is the cross-cut of all maximal

[9] G. Frobenius, *Über endliche Gruppen*, Sitzungsber. Akad. Berlin, 1895 (I), pp. 163–194.
[10] Ore II, Chapter 2.

normal subgroups of G, then $C = C_1 \cap \cdots \cap C_n = M_1 \cup \cdots \cup M_n$. The quotient group

$$L = G/C \tag{3}$$

has been called the *upper normal cover quotient* of G. It is completely reducible; i.e., it is the direct product of simple groups $L = R_1 \cup \cdots \cup R_n$, where $R_i = N_i/C$, $N_i = M_1 \cup \cdots \cup M_{i-1} \cup A_i \cup M_{i+1} \cup \cdots \cup M_n$, and one has the isomorphisms $R_i \cong G_i \cong P_i$. This shows that except for isomorphisms the simple quotient groups P_i to which the components A_i belong are uniquely determined.

2. **Groups whose maximal subgroups are normal.** We shall now study the groups in which all maximal groups are permutable. We have already seen (Theorem 5, Chapter I) that this implies that all maximal groups are normal. In this case all the maximal groups must have quotient groups which are cyclic of prime order. This implies in turn that all the indecomposable A_i in a normal decomposition (1) must be of Abelian type, and hence the upper normal cover quotient (3) is Abelian and the direct product of cyclic groups of prime order.

We must also have $C = \phi$, where ϕ is the ϕ-group of G, since C was the cross-cut of all normal maximal groups while ϕ is the cross-cut of all maximal groups in G. The solution of our problem then follows from a theorem of Wielandt:[11]

The necessary and sufficient condition for a group to be nilpotent, i.e., the direct product of its Sylow groups, is that the ϕ-group contain the commutator group.

This leads immediately to the following characterization of a nilpotent group:

THEOREM 1. *The necessary and sufficient condition for a group to be nilpotent is that all maximal subgroups be permutable, hence that they all should be normal.*

Proof. It is easily seen that every nilpotent group has this property since it holds for the Sylow groups. The converse follows from the fact that $G/C = G/\phi$ is Abelian, and hence ϕ contains the commutator group.

3. **Cross-cut decompositions.** To the theory of decomposition given in §1 there corresponds a dual theory which one obtains by considering the quotient groups G/A for normal subgroups A and introducing the operations

$$G/A \cup G/B = G/A \cap B, \qquad G/A \cap G/B = G/A \cup B.$$

The dual of the decomposition (1) becomes the representation

$$A_1 \cap A_2 \cap \cdots \cap A_n = E \tag{4}$$

of the unit group as the cross-cut of normal subgroups, where the A_i are normally cross-cut indecomposable in G; i.e., they cannot be represented as the cross-cut of larger normal subgroups of G. The necessary and sufficient con-

[11] H. Wielandt, *Eine Kennzeichnung der direkten Producte von p-Gruppen*, Math. Zeit., vol. 41(1936), pp. 281–282.

dition for A_i to have this property is that among the various normal subgroups of G containing A_i as a proper subgroup there exists a unique minimal one M_i. We say again that A_i belongs to the quotient group $M_i/A_i = P_i$. When the A_i in (4) are indecomposable in the cross-cut sense, none of them can be replaced by any larger normal group.

Now let D denote the (lower) normal *cover group* of G, i.e., the union of all minimal normal subgroups of G. Since the decomposition (4) is supposed not to contain any superfluous components, A_i cannot contain D, because D has a normal subgroup in common with any normal subgroup of G. But since there exists a unique minimal M_i containing A_i, one must have $A_i \cup \bar{P}_i = M_i$ for every minimal normal subgroup $\bar{P}_i$ of G not contained in A_i. This shows that the P_i to which A_i belongs is isomorphic to a minimal normal subgroup of G. One also sees that A_i contains all but one minimal group in a suitable representation of D as the direct product of minimal normal subgroups of G, and the A_i can be defined as maximal groups with this property.

It follows by duality that any two representations (4) have the same number of components, the minimal groups corresponding to these components are isomorphic in some order, and one component in one representation can always replace a suitably chosen component in the other. There are also certain cases in which the A_i must be unique, analogously to the dual cases.[12]

4. **Representations by permutations.** The preceding theory is closely connected with the problem of representing G as a permutation group. Let us recall that any transitive homomorphic representation of G may be obtained by expanding G in co-sets with respect to a subgroup A,

$$G = \sum Ag_i, \tag{5}$$

and when the co-sets are multiplied on the right by the elements of G, one obtains a set of permutations of the co-sets forming a group P_A homomorphic to G and isomorphic to G/N, where N is the greatest normal subgroup of G contained in A. We shall say that the representation P_A is *induced* by A, and it is convenient to say that A and also P_A *belong* to N. The *degree* of the transitive representation P_A is k, where k is the index of A in G.

Any intransitive representation of G may be obtained by aligning or forming the sum of transitive representations. Any representation has a corresponding set of subgroups

$$A_1, A_2, \cdots, A_s \tag{6}$$

inducing the transitive components and each of them belongs to a normal subgroup of G

$$N_1, N_2, \cdots, N_s \tag{7}$$

[12] This short exposition of the properties of cross-cut decompositions is mainly a restatement from Ore II, Chapter 2, §6. It has been repeated here because the presentation at one point was not quite clear.

and the degree of the representation is $k_1 + \cdots + k_s$, where k_i is the index of A_i in G.

THEOREM 2. *The necessary and sufficient condition for a permutation representation of G induced by the subgroups A_i in (6) to be a true representation is that the normal subgroups N_i in (7) to which the A_i belong satisfy the relation*

$$N_1 \cap \cdots \cap N_s = E. \tag{8}$$

Proof. An element corresponding to the unit permutation must belong to all N_i, and conversely.

All transitive representations belonging to the same normal subgroup N shall be said to belong to the same class C_N. In each class there is contained a regular permutation representation induced by N itself. Any group A containing N corresponds to a set of systems of imprimitivity in this regular representation, and any normal group N_1 is characterized by the property that its elements transform the systems of imprimitivity into themselves. The quotient group G/N_1 is then isomorphic to the permutations of the systems.

The classes of permutation representation can be made into a structure by the definitions

$$C_N \cup C_M = C_{N \cup M}, \qquad C_N \cap C_M = C_{N \cap M}. \tag{9}$$

We shall say that a class C_N is indecomposable if there exist no greater normal subgroups L and M such that $N = M \cap L$. If N is decomposable, the group G/N can be represented isomorphically as the sum of a representation of G/M and G/L.

We shall now discuss the true or isomorphic representations of G as a permutation group. We shall usually suppose that the representation is reduced; i.e., no N_i in (8) shall contain the cross-cut of any set of the others. When such N_i are omitted, the relation (8) still holds. In terms of permutations this reduction means that one omits components homomorphic to a sum of the others.

From (9) follows that the structure of representative classes is isomorphic to the structure of all normal subgroups of G. When this is combined with Theorem 2, the decomposition theory indicated in §3 gives

THEOREM 3. *Any true reduced representation of a group as a permutation group is the sum of transitive indecomposable components. Any two such representations contain the same number of transitive components and one component in one representation may be exchanged with a suitably chosen component in the other.*

All the indecomposable components may be obtained by constructing the representation classes corresponding to the various subgroups A_i occurring in a reduced decomposition (4). The number of transitive indecomposable components in any permutation representation is equal to the number r of independent minimal normal groups of G in the representation of the cover D as the

direct product

$$D = \bar{P}_1 \cup \bar{P}_2 \cup \cdots \cup \bar{P}_r .$$

For many problems it is of importance to determine the true representations of smallest degree. Such a representation must be reduced. Furthermore, in each class C_N one must select the representation of smallest degree, i.e., the representation induced by a subgroup A belonging to N and having the smallest index in G. The effect of decomposing a class is usually to obtain smaller representations, and it is to be conjectured that a sum of indecomposable representations gives the absolute smallest degree.

For Abelian groups the problem is readily solved. Here there is only one representation, namely, the regular one, in each class. If a class is decomposed $N = M \cup L$ with the indices k_N, k_M, and k_L, one obtains a shorter representation of degree $k_M + k_L < k_N$. This shows that the shortest representation must be a sum of cyclic transitive parts and one finds without difficulty

THEOREM 4. *The degree of the smallest true representation of an Abelian group as a permutation group is equal to the sum of its invariants.*[13]

5. **Applications.** If a group G contains a subgroup H, then any true representation of G contains a true representation of H. The transitive representations of G are induced by subgroups A not containing any normal subgroup of G. If d_H is the degree of a smallest representation of H, then it follows that the index of any such A must be at least d_H. From Theorem 4 follows that if G contains an Abelian group H, then the index of any subgroup A of G not containing normal subgroups is at least equal to the sum of the invariants of H. As a special case we have

THEOREM 5. *Let G be a simple group and H a maximal Abelian subgroup of G with the invariants $p_i^{\alpha_i}$. Then the index of any maximal group in G is at least $\sum_i p_i^{\alpha_i}$.*

If one could give some lower bound for the degree of a true permutation representation of a p-group, this would also give a lower bound for the indices of subgroups in a simple group from the group order.

Now let A and B be permutable groups and let

$$B = \sum b_i D \tag{10}$$

be the co-set expansion of B with respect to their cross-cut. We form all products

$$a \cdot b_i = b_i' \cdot a_i' \tag{11}$$

[13] A. Powsner, *Über eine Substitutionsgruppe kleinsten Grades die einer gegebenen Abelschen Gruppe isomorph ist.* After the completion of this paper a review of this paper appeared (Zentralblatt f. Math., vol. 19(1939), pp. 155–156) stating that it contained a theorem equivalent to Theorem 4.

for a fixed element a in A and all $\frac{n_B}{n_D}$ generators b_i in (10). The b_i' also denote some of these generators. No two of them can be equal since $b_i' = b_j'$ implies $ab_i a_i'^{-1} = ab_j a_j'^{-1}$ or $b_j^{-1} \cdot b_i \subset A$ and b_i and b_j would belong to the same co-set in (10).

Each element a in A corresponds therefore to a permutation of the generators g_i defined by the relation (11). This permutation can be assumed to be a substitution on at most $\frac{n_B}{n_D} - 1$ letters since one can always choose $b_1 = e$.

The group A is homomorphic to the substitution group thus defined. Those elements a_0 in A which correspond to the unit substitution form a subgroup A_0 for which $bA_0b^{-1} \subset A$ for every b in B. According to Theorem 11, Chapter I, this implies that A_0 is normal in $A \cup B$.

Such a normal subgroup $A_0 \supset E$ must exist if A or a subgroup of A cannot be represented isomorphically by a permutation group on $\frac{n_B}{n_D} - 1$ letters. Let us mention only one case.

THEOREM 6. *Let A and B be permutable groups. If A contains any element of prime power order p^α such that $p^\alpha \geqq \frac{n_B}{n_D}$, then A must contain a normal subgroup of $A \cup B$.*

6. **The ϕ-group.** The ϕ-group of a group G is a characteristic subgroup defined as the cross-cut of all maximal subgroups of G. It may also be defined as the set of elements which can be omitted in any generating system for G. It is also the union of all those subgroups which are superfluous in any representation of G as the union of subgroups. A fundamental property of the ϕ-group is the *basis theorem*[14] expressing that ϕ is the maximal normal subgroup such that if $g_1, \cdots, g_s$ is any generating system of the co-sets in G/ϕ, then they also generate G. A special case is Burnside's basis theorem for p-groups. From the basis theorem follows that the ϕ-group of the quotient group G/ϕ is the unit group, so that one cannot form repeated ascending ϕ-groups.

If A is a normal subgroup of G, then

$$\phi_A \subseteq A \cap \phi_G . \tag{12}$$

This follows from the fact that any generating system for G/ϕ_A must also generate A/ϕ_A. From (12) one concludes $\phi_{A \cap B} \subseteq \phi_A \cap \phi_B$. Similarly, one obtains $\phi_{A \cup B} \supseteq \phi_A \cup \phi_B$. If A and B are relatively prime, one finds that the equality holds in the last relation.

The ϕ-group is nilpotent.[15] This fact may be considered as a special case of the following

[14] H. Zassenhaus, *Lehrbuch der Gruppentheorie*, vol. 1, p. 45.

[15] See, for example, Miller, Blichfeldt, and Dickson, *Theory and Application of Finite Groups*, pp. 71–72.

THEOREM 7. *Let A be a normal subgroup of G and N_p the normalizer in G of some Sylow group S_p of A. Then*

$$G = N_p \cup A. \tag{13}$$

Proof. All the conjugates of S_p belong to A and therefore have the form aS_pa^{-1}. This means that for any g in G there exists an a such that $gS_pg^{-1} = aS_pa^{-1}$, or $g = a \cdot n_p$, where n_p belongs to N_p.

When A is the ϕ-group, (13) reduces to $G = N_p$ and S_p is normal.

The ϕ-group may be determined by means of this property. We observe first that any group has a maximal normal nilpotent subgroup N. This group may be obtained as follows: The union of all normal p-groups of G for any prime p is a maximal normal p-group N_p and the group N is given by $N = N_p \cup N_q \cup \cdots$ for the various primes $p, q, \cdots$ dividing the group order.

Since ϕ is nilpotent, we can conclude by means of (12) that

$$G \supset N \supset \phi_G \supset \phi_N. \tag{14}$$

It remains therefore to determine the location of ϕ_G between N and ϕ_N. In a nilpotent group the quotient group N/ϕ_N is Abelian and the union of groups of the type $(p, p, \cdots)$ for the various primes $p, q, \cdots$. In the following it is convenient to say that the group G *splits* over the normal subgroup A if there exists a subgroup M such that

$$G = M \cup A, \qquad M \cap A = B, \tag{15}$$

where B also is normal in G.

Now let us consider the quotient group G/ϕ_N. From the definition of the ϕ-group it follows that G/ϕ_N cannot split over ϕ/ϕ_N. But for any normal subgroup A not contained in ϕ such that $N \supset A \supset \phi_N$ the groups G and G/ϕ_N must split since for any maximal group M of G not containing A we must have the relations (15). Here B is normal in G since it is normal in M and also normal in A since the quotient group A/ϕ_N is Abelian. We can summarize these remarks as follows:

THEOREM 8. *Let G be a finite group, N its maximal normal nilpotent group and ϕ and ϕ_N the ϕ-groups of G and N. Then $N \supseteqq \phi \supseteqq \phi_N$ and ϕ/ϕ_N is the maximal normal subgroup of G/ϕ_N contained in the Abelian group N/ϕ_N over which G does not split.*

This result may also be expressed as follows:

THEOREM 9. *Let G be a finite group. The necessary and sufficient condition for G not to have a ϕ-group is that the maximal normal nilpotent subgroup of G be the unit group or the union of Abelian groups of type $(p, p, \cdots)$ and in the last case G must split over all its minimal normal Abelian subgroups.*

Chapter III. Properties of chains

1. **Refinement of chains.** A set of subgroups

$$G \supset A_1 \supset \cdots \supset A_{r-1} \supset E \tag{1}$$

shall be called a *chain* and the indices $[A_{i-1} : A_i]$ are the *indices of the chain*. Two chains are said to have the same *index type* or to be *conformal* when their indices are the same in some order. When each group in (1) is maximal in the preceding, the chain is said to be *complete* and the groups $A_{i-1} \supset A_i$ are said to be *consecutive*.

As usual we call (1) a *composition chain* when A_i is a maximal normal subgroup of A_{i-1} and a *principal chain* when A_i is maximal among the normal subgroups of G contained in A_{i-1}. Similarly, by introducing the concept of quasi-normality one can define *quasi-principal chains* while a quasi-composition chain is a composition chain according to Theorem 16, Chapter I.

In the following we shall compare two chains

$$G = A_0 \supset A_1 \supset \cdots \supset A_{r-1} \supset A_r = E,$$
$$G = B_0 \supset B_1 \supset \cdots \supset B_{s-1} \supset B_s = E. \tag{2}$$

This will be done by means of a weak form of the theorem of Jordan-Hölder which we have previously derived.[16] Let us recall first a few of the most important facts.

The two chains (2) are said to be *cross-cut permutable* when A_i $(i = 0, 1, \cdots, r)$ is permutable with all groups $A_{i-1} \cap B_j$ $(j = 0, 1, \cdots, s)$ and B_j $(j = 0, 1, \cdots, s)$ is permutable with all $B_{j-1} \cap A_i$ $(i = 0, 1, \cdots, r)$. The main theorem on such chains is:

Any two cross-cut permutable chains may be refined into two new cross-cut permutable chains which are conformal.

The refined chains may be given explicitly. They consist of the terms

$$A_{i,j} = A_i \cup (A_{i-1} \cap B_j), \qquad B_{k,l} = B_k \cup (B_{k-1} \cap A_l) \tag{3}$$

and one has $[A_{i,j-1} : A_{i,j}] = [B_{j,i-1} : B_{j,i}]$.

To this theory there exists a dual: We say that two chains (2) are *union permutable* when A_{i-1} is permutable with $A_i \cup B_j$ $(j = 0, 1, \cdots, s)$ and B_{j-1} is permutable with $A_i \cup B_j$ $(i = 0, 1, \cdots, r)$. For such chains one has the same refinement theorem as above. In this case the refinements corresponding to (3) are

$$\bar{A}_{i,j} = A_{i-1} \cap (A_i \cup B_j), \qquad \bar{B}_{k,l} = B_{k-1} \cup (B_k \cap A_l). \tag{4}$$

2. **Comparison with principal chains.** We shall now apply the preceding results to the case where the second chain in (2) is a principal or quasi-principal chain. Then obviously B_i is permutable with $B_{i-1} \cap A_j$ and B_{i-1} is permutable with $B_i \cup A_j$. Furthermore, for any elements a_i in A_i and b_j in B_j we have $a_i \cdot b_j = b_j' \cdot a_i^n$, and if b_j belongs to $A_{i-1} \cap B_j$, then b_j' must have the same property. This shows that A_i and $A_{i-1} \cap B_j$ are permutable and similarly one sees that A_{i-1} is permutable with $A_i \cup B_j$. This proves

[16] Ore I, Chapter 3.

THEOREM 1. *Any chain is both cross-cut and union permutable with a principal or quasi-principal chain.*

In this case the preceding theory is somewhat simplified since it follows from the Dedekind relation that the refinements (3) and (4) are the same.

From now on we shall study the case where the first chain in (2) is a complete chain while the second is a principal or quasi-principal chain. Then the second chain can be refined into a chain which has the same index type as the first and we have

THEOREM 2. *To any complete chain there exists a conformal chain (not necessarily complete) passing through any prescribed principal or quasi-principal chain.*

We conclude from this theorem

THEOREM 3. *A complete chain is never shorter than a principal or quasi-principal chain.*

A further important consequence of Theorem 2 is

THEOREM 4. *Let n_i $(i = 1, 2, \cdots, s)$ denote the indices of any principal or quasi-principal chain. It is then possible to enumerate the indices $f_j^{(i)}$ of any complete chain in such a manner that $n_i = f_1^{(i)} f_2^{(i)} \cdots f_{r_i}^{(i)}$ $(i = 1, 2, \cdots, s)$.*

Another point of interest about the refinement into a principal or quasi-principal chain is the following: If a group A_i in the first chain is normal in A_{i-1} and the second chain is a principal chain, then $B_{j,i}$ is found to be normal in $B_{j,i-1}$. Hence when the first chain is complete and contains a certain number of cases in which one term is normal in the preceding, then the refinement into a principal chain must contain at least the same number of normalities.

3. **Chains of maximal length.** Now let us turn to the case where the first chain in (2) is a chain of maximal length in G while the second remains principal or quasi-principal. In this case the refinement of the first into the second must also be complete and we have therefore

THEOREM 5. *To any complete chain of maximal length there always exists a conformal chain containing any prescribed principal or quasi-principal chain.*

Theorem 5 shows that one can determine all index types of maximal chains in a group from those of the quotient groups B_{j-1}/B_j in any principal chain. It is possible to reduce the problem still a little further. It is known that the group B_{j-1}/B_j is the direct product of simple isomorphic groups and hence there exists a principal chain in the quotient group

$$B_{j-1} \supset B_{j,1} \supset \cdots \supset B_{j,k} \supset B_j, \tag{5}$$

where all quotient groups are simple, and hence the $B_{j,i}$ form part of a composition chain in G. When Theorem 5 is applied to the quotient group B_{j-1}/B_j, one sees that to any maximal chain in it there exists another with the same index

type passing through all $B_{j,i}$. By means of the theorem of Jordan-Hölder we obtain therefore

THEOREM 6. *To any chain of maximal length* $G \supset A_1 \supset \cdots \supset A_{r-1} \supset E$ *there exists a conformal chain passing through any prescribed composition chain* $G \supset B_1 \supset \cdots \supset B_{s-1} \supset E$. *Hence all index types of maximal chains in* G *may be obtained by stringing together all such types of the simple quotient groups in a composition chain.*

This theorem shows that if m_j is the maximal length of chains in any simple group B_{j-1}/B_j, then the maximal length of any chain in G is $m_1 + m_2 + \cdots + m_s$.

Let us observe finally

THEOREM 7. *The necessary and sufficient condition that all complete chains of maximal length shall be conformal is that all the simple quotient groups in a composition chain have this property.*

Chapter IV. Solvable groups

1. **Properties of chains.** We shall now apply the preceding results to derive various properties of the solvable groups. We shall study in particular the properties of maximal groups in solvable groups.

A solvable group may be defined as a group in which there exists a composition chain whose indices are primes. From this definition follows trivially

THEOREM 1. *In a solvable group all chains of maximal length are conformal.*

The same result must obviously hold in any group in which there exists a chain whose indices are all primes.

The solvable groups also have the characteristic property that there exists a principal chain

$$G \supset L_1 \supset \cdots \supset L_{r-1} \supset E, \tag{1}$$

where the quotient groups L_{i-1}/L_i are Abelian of type $(p, p, \cdots)$. There even exist characteristic chains with this property. When this fact is combined with Theorem 4, Chapter III, one obtains

THEOREM 2. *In a solvable group the index of any two consecutive subgroups is a prime power dividing one of the indices in a principal chain.*

From Theorem 9, Chapter I one concludes further:

THEOREM 3. *In a pair of consecutive groups* $A \supset B$ *in a solvable group,* B *is permutably contained in* A.

2. **Decompositions of groups.** Let H be a subgroup of some group G. It is convenient to say that H is *decomposably contained* in G if there exist two normal subgroups $N' \supset N$ in G such that

$$H \cup N' = G, \qquad H \cap N' = N. \tag{2}$$

We shall call N' the *normal component* of the decomposition (2) and say as before that G *splits* over N'/N. We say further that G *splits regularly* over N'/N if any other decomposition $H_1 \cup N' = G$, $H_1 \cap N' = N$ implies that H and H_1 are conjugate in G.

We shall now prove various theorems which are of importance for the following.

THEOREM 4. *Let M be a maximal subgroup in an arbitrary finite group G and let us suppose that G contains some minimal Abelian normal subgroup L of order p^α. If L is not contained in M, then M is decomposably contained in G with normal component L*

$$M \cup L = G, \qquad M \cap L = E \tag{3}$$

and M is either normal or has p^α conjugates.

Proof. The first relation (3) is obvious when L is not contained in M. The second follows from the fact that $M \cap L$ is normal in M and also in L since L is Abelian, and consequently $M \cap L$ is normal in G and therefore equal to E. The last statement in Theorem 4 follows from

THEOREM 5. *A maximal group is normal or has as many conjugates as its index.*

Proof. If M is not normal in G, it is its own normalizer.

It should also be noted in connection with Theorem 4 that if M in (3) does not contain any normal subgroup of G, the normal subgroup L is uniquely determined. If, namely, $G = M \cup L_1$, $M \cap L_1 = E$ were another decomposition, where L_1 also is normal in G, then one sees that L_1 must be minimal normal and Abelian and one finds by the Dedekind relation $L \cup L_1 = L \cup (M \cap (L \cup L_1))$. Here $M \cap (L \cup L_1)$ is normal in M and also in $L \cup L_1$ since this group is Abelian, and hence $M \cap (L \cup L_1)$ would be a normal subgroup of G contained in M.

THEOREM 6. *Let B be normal in A and let A split regularly over B*

$$A = H \cup B, \qquad B \cap H = E \tag{4}$$

such that H is its own normalizer in A. Then any group G in which A and B are contained normally will also split regularly

$$G = N \cup B, \qquad N \cap B = E, \tag{5}$$

where N is the normalizer of H and also its own normalizer in G.

Proof. Since all conjugates of H belong to A, we have for any g in G, $gHg^{-1} = aHa^{-1}$, where a belongs to A. This gives $g = a \cdot n$, where n belongs to the normalizer N of H in G and the first relation (5) is obtained. Since none of the elements of B can belong to N, the second follows.

Now let us suppose that there exists some relation (5) for a subgroup N of G. Then one finds $A = B \cup (N \cap A)$, $(N \cap A) \cap B = E$ and $N \cap A$ must be a conjugate of H. Since $N \cap A$ is normal in N, it follows that N must be the

normalizer of this conjugate of H and the decomposition (5) is regular. If any element of B transformed N into itself, then N would contain more than one conjugate of H, and this is impossible.

Theorem 6 may be used to establish the existence of certain maximal groups. We shall apply it in the case where G contains the normal subgroups $A \supset B$, where we suppose that B is a unique minimal normal Abelian subgroup of order p^α and type $(p, p, \cdots)$. Furthermore A/B shall be a minimal normal Abelian subgroup of G/B of order q^β and type $(q, q, \cdots)$, where $p \neq q$. Then A splits regularly over B with $A = C \cup B$, $C \cap B = E$, where C is a Sylow group corresponding to q. Let us show further that C is its own normalizer in A. We denote by B_1 the subgroup of B consisting of those elements which belong to the normalizer of C. Since $B_1 = B \cap (C \cup B_1)$, it follows that B_1 is normal in $C \cup B_1$, and hence the elements of B_1 and C are permutable and B_1 belongs to the center of A. This center must be normal in G, so we have $B_1 = B$ or $B_1 = E$. The first possibility is excluded since C would be a characteristic subgroup of A contrary to the assumption that B was the only minimal normal Abelian subgroup of G. It also follows that the number of conjugates of C is p^α, and hence $p^\alpha \equiv 1 \pmod q$ since C is a Sylow group.

We conclude from Theorem 6

THEOREM 7. *Let G contain the normal subgroups $A \supset B$, where B is a unique minimal normal Abelian subgroup of G of order p^α and type $(p, p, \cdots)$, while A/B is some minimal normal Abelian subgroup of G/B of order q^β and type $(q, q, \cdots)$ with $q \neq p$. Then A splits regularly*

$$A = C \cup B, \qquad C \cap B = E,$$

where C is isomorphic to A/B and is its own normalizer in A. The group G also splits regularly

$$G = M \cup B, \qquad M \cap B = E,$$

where M is the normalizer of C in G. Furthermore, M is maximal in G and the number of its conjugates is $k = p^\alpha \equiv 1 \pmod q$.

Proof. It remains only to show that M is maximal in G. If for some maximal subgroup M_1 one has $M_1 \supset M$, then $M_1 = M \cup (M_1 \cap B)$ and it follows from the proof of Theorem 4 that $M_1 \cap B = B$ or $M_1 \cap B = E$, and this gives $M_1 = G$ or $M_1 = M$.

3. **Properties of solvable and nilpotent groups.** It follows already from §1 that in a solvable group G any maximal group M is permutably contained and its index is a power of a prime. This result may be improved as

THEOREM 8. *Let G be solvable and M any maximal group. Then M is normally or decomposably contained in G and the index of M is equal to an index in a principal chain.*

Proof. We shall say as before that a group M in G *belongs* to the normal subgroup N of G if N is the maximal normal subgroup of G which M contains. Now let M be maximal. In the group G/N the maximal group M/N must belong to the unit group. Let L/N be a minimal normal subgroup of G/N of order p^α. From Theorem 4 it follows that $M \cup L = G, L \cap M = N$, and our theorem is proved.

From Theorem 8 one obtains the following characterization of solvable groups:

THEOREM 9. *The necessary and sufficient condition for a group to be solvable is that in any complete chain any group shall always be normally or decomposably contained in the preceding.*

Proof. Theorem 8 shows that every chain in a solvable group must have this property. Conversely, when every chain has this property, G must contain normal subgroups, and the theorem follows by induction.

Theorem 9 may be considered an analogue of the following characterization of nilpotent groups:

THEOREM 10. *The necessary and sufficient condition for a group to be nilpotent is that any complete chain be a composition chain.*

It is easily seen that every nilpotent group has this property. The converse follows from the fact that this condition on the chains implies that no subgroup can be its own normalizer, while in any group the normalizers of the Sylow groups are their own normalizers.

4. **Maximal groups.** We shall now turn to the special properties of maximal subgroups in a solvable group G. We prove first

THEOREM 11. *Let G be solvable and N a normal subgroup. All maximal subgroups of G belonging to N are conjugate, and conversely, all conjugate maximal subgroups belong to the same N.*

Proof. The last part of the theorem is obvious. To prove the first let us observe that there is no limitation in assuming $N = E$ since otherwise one need only consider the quotient groups. Then M is a maximal group which does not contain any normal subgroup of G, and hence according to Theorem 4 there exists a unique minimal normal subgroup L of order p^α such that $M \cup L = G$, $M \cap L = E$. Now let A/L be a minimal normal subgroup of G/L of order q^β.

We shall first have to show that $p \neq q$. Suppose, namely, that $q = p$. Then A as a p-group must have a center. On the other hand, one has $A = L \cup (M \cap A)$. Since the center of A is normal in G, it follows that L must belong to it, and this implies that $M \cap A$ is normal in G contrary to the assumption.

We have therefore $q \neq p$ and $A = C \cup L, C \cap L = E$, where C is a Sylow group of A corresponding to q. The conditions of Theorem 7 are satisfied and all maximal groups of G not containing L must be normalizers of C or of one of its conjugates.

This proof permits us also to state the conditions for the existence of a maximal group corresponding to a given N.

THEOREM 12. *Let N be a normal subgroup of the solvable group G. The necessary and sufficient conditions for the existence of maximal subgroups of G belonging to N are*:

(1) *N must be normally cross-cut indecomposable in G. Then there exists a single minimal normal subgroup N_1/N of G/N of order p^α.*

(2) *The group G/N_1 shall not contain any normal subgroup whose order is a power of p.*

Let us prove next

THEOREM 13. *Let G be a solvable group. To every prime p dividing the group order there exists a maximal group whose index is a power of p.*

Proof. This theorem is a consequence of Theorem 12. It may also be proved independently by induction with respect to the group order. Let B be a minimal normal subgroup. Since the theorem holds in G/B, we need only consider the case where the order of B is p^α, while the order of G/B is not divisible by p. Furthermore, one can assume that B is the only minimal normal subgroup of G. Now let A/B be a minimal normal subgroup of G/B. Then A/B is Abelian of order q^β and the conditions of Theorem 7 are satisfied. If C is a Sylow group corresponding to q in A, then its normalizer M in G is a maximal subgroup with the index p^α.

By repeated applications of Theorem 13 one obtains the following theorem of P. Hall:[17]

Let G be a solvable group of order $N = n \cdot m$, where n and m are relatively prime. Then G has subgroups of orders n and m.

Hall[18] has also shown that any group with this property must be solvable. This implies that every group in which every subgroup has the property of Theorem 13 must be solvable.

Hall has also obtained the following results:

Any two subgroups of order m are conjugate.

Any subgroup of G whose order divides m belongs to some subgroup of order m.

The number of subgroups of order m is a product of prime power factors p^α, each dividing an index in a principal chain and $p^\alpha \equiv 1 \pmod q$ for some prime factor q of m.

We shall conclude our investigations on maximal groups in solvable groups by proving

THEOREM 14. *In a solvable group any two maximal groups are permutable or conjugate.*

[17] P. Hall, *A note on soluble groups*, Journal of the London Math. Soc., vol. 3(1928), pp. 98–105.

[18] P. Hall, *A characteristic property of soluble groups*, Journ. London Math. Soc., vol. 12(1937), pp. 198–200.

Proof. We shall prove the theorem by induction with respect to the group order. Let M_1 and M_2 be a pair of maximal groups in G. If both M_1 and M_2 contain a minimal normal subgroup L of G, then M_1 and M_2 are conjugate or permutable since M_1/L and M_2/L have this property. We may assume therefore that M_1 and M_2 have no common minimal normal subgroup of G. Now let us suppose that M_1 contains such a group L. Then according to Theorem 4 we have $G = M_2 \cup L$, $M_2 \cap L = E$, and consequently, $M_1 = L \cup (M_1 \cap M_2)$. This representation of M_1 shows, however, that it is permutable with M_2. In the final case where neither M_1 nor M_2 contain any normal subgroup of G, they both belong to the unit group E and are therefore conjugate according to Theorem 11.

There are two extreme cases in Theorem 14. When all maximal groups are permutable, then we have already seen that the group is nilpotent. The case where all maximal subgroups are conjugate is solved by the following theorem.

THEOREM 15. *A group in which all maximal subgroups are conjugate is cyclic.*

Proof. Let $M_1, M_2, \cdots$ be the maximal groups. Since it is known that the elements belonging to these groups cannot be all elements of G, there must exist an element g not in any M_i and hence g must generate G.

This theorem may be considered a generalization of the well-known theorem that a p-group of order p^α which contains only one subgroup of index p must be cyclic.

5. **Chains in solvable groups.** In all finite groups there exist *normal completely reducible* chains $G = C_l \supset C_{l-1} \supset \cdots \supset C_2 \supset C_1 \supset E$, where C_{i-1}/C_i is the union of minimal normal subgroups of G/C_i. In a solvable group the quotient groups C_{i-1}/C_i are Abelian groups in which every element has prime order. For these chains various interesting properties have been derived.

In a solvable group one can also construct *normal nilpotent* chains

$$G = N_k \supset N_{k-1} \supset \cdots \supset N_2 \supset N_1 \supset E, \tag{6}$$

where each N_{i-1}/N_i is a normal nilpotent subgroup of G/N_i. We have already observed that in any group there exists a maximal normal nilpotent subgroup M_1. Similarly, G/M_1 has a maximal nilpotent normal subgroup M_2/M_1, and by continuing this process, one obtains in a solvable group a unique maximal normal nilpotent ascending chain

$$G = M_h \supset M_{h-1} \supset \cdots \supset M_2 \supset M_1 \supset E. \tag{7}$$

One can then prove the following:

Let (7) *be the maximal normal nilpotent ascending chain in a solvable group G while* (6) *is any normal nilpotent chain. Then for every i $M_i \supset N_i$.*

From this result follows:

The maximal nilpotent chain has the shortest length among the normal nilpotent chains.

These theorems are quite analogous to the results on normal completely reducible chains and various other theorems also hold for both classes of chains.

Let us now dualize the preceding theory. In any group there exists a smallest normal subgroup $\bar{M}_p$ such that $G/\bar{M}_p$ is a p-group. The cross-cut of all $\bar{M}_p$ gives the smallest normal subgroup $\bar{M}$ such that $G/\bar{M}$ is a nilpotent quotient group. Obviously $\bar{M}$ is a characteristic subgroup. In $\bar{M}$ there exists a corresponding smallest normal group with nilpotent quotient group. Thus one obtains in any solvable group a maximal normal nilpotent descending chain

$$G = \bar{M}_h \supset \bar{M}_{h-1} \supset \cdots \supset \bar{M}_2 \supset \bar{M}_1 \supset E. \tag{8}$$

It follows then further:

The descending and ascending maximal normal nilpotent chains (8) *and* (7) *in a solvable group both have the same length and are shortest normal nilpotent chains. If* $G = N_h \supset N_{h-1} \supset \cdots \supset N_2 \supset N_1 \supset E$ *is any shortest normal nilpotent chain in* G, *then one has for every* i $M_i \supset N_i \supset \bar{M}_i$.

One can also consider normal Abelian chains

$$G = A_0 \supset A_1 \supset \cdots \supset A_k \supset E, \tag{9}$$

where every quotient group A_{i-1}/A_i is Abelian. Such chains must exist in every solvable group. But here one cannot usually define the maximal normal Abelian subgroup since the union of two normal Abelian groups need not be Abelian. One could avoid this difficulty by considering groups which are the union of Abelian normal subgroups. There exists, however, a smallest characteristic group with Abelian quotient group, namely, the commutator group. Correspondingly, one obtains the chain of commutator groups

$$G = C_0 \supset C_1 \supset \cdots \supset C_m \supset E, \tag{10}$$

and we can say: *The commutator chain* (10) *is a shortest chain among the normal Abelian chains, and if* (10) *is the commutator chain and* (9) *an arbitrary Abelian chain, then for every* i $A_i \supset C_i$.

Chapter V. Groups with conformal chains

1. **Dispersible groups.** We shall now study various classes of special solvable groups.

Let N be the order of a group G and let

$$N = p_1^{\alpha_1} p_2^{\alpha_2} \cdots p_r^{\alpha_r}, \qquad p_1 > p_2 > \cdots > p_r \tag{1}$$

be its prime factorization with the prime factors arranged in decreasing order. A type of group which occurs in various investigations is that in which there exists a chain of normal subgroups in G

$$E \subset K_1 \subset K_2 \subset \cdots \subset K_r = G \tag{2}$$

with the corresponding orders

$$1,\ p_1^{\alpha_1},\ p_1^{\alpha_1} p_2^{\alpha_2},\ \cdots,\ N. \tag{3}$$

Such groups shall be called *dispersible groups*. In certain cases one can also introduce the group orders (3) in reverse order of the primes

$$1,\ p_r^{\alpha_r},\ p_r^{\alpha_r} p_{r-1}^{\alpha_{r-1}},\ \cdots,\ N. \tag{4}$$

We shall say then that G is a *reverse dispersible group*.

Both types of dispersible groups are obviously solvable and form a type of group including the nilpotent groups. It follows from the results of Hall mentioned above that if $K_i \supset K_j$ in a chain (2), then K_i splits regularly over K_j. The chain (2) is also a characteristic chain of the group.

Now let G be any group. It is easily seen that the union of two dispersible groups is again dispersible, and hence there exists a unique maximal normal dispersible subgroup. This shows the existence of an ascending chain of maximal normal dispersible groups

$$G = D_k \supset D_{k-1} \supset \cdots \supset D_2 \supset D_1 \supset E \tag{5}$$

in any solvable group. These groups are all characteristic subgroups. Similarly, one sees the existence of a minimal normal group $\bar{D}$ such that $G/\bar{D}$ is a dispersible group and one obtains also a descending chain of maximal normal dispersible groups

$$G = \bar{D}_k \supset \bar{D}_{k-1} \supset \cdots \supset \bar{D}_2 \supset \bar{D}_1 \supset E. \tag{6}$$

As before one finds that both chains (5) and (6) have the same length and are shortest chains among all normal dispersible chains. If $G = C_k \supset C_{k-1} \supset \cdots \supset C_2 \supset C_1 \supset E$ is any shortest dispersible chain, then one finds as before that $D_i \supset C_i \supset \bar{D}_i$.

Among other properties let us mention that any subgroup and any quotient group of a dispersible group is again dispersible.

The following theorem does not hold for reverse dispersible groups:

THEOREM 1. *The necessary and sufficient condition for a group to be dispersible is that there exist a complete chain $E \subset A_1 \subset A_2 \subset \cdots \subset A_t = G$, where the indices are primes in non-increasing order.*

Proof. The theorem follows by induction with respect to the group order. It is then sufficient to prove that G contains a normal subgroup of order $p_1^{\alpha_1}$ because one can apply the same argument to the quotient groups. Now according to assumption A_{t-1} contains a normal subgroup of order $p_1^{\alpha_1}$ and this subgroup must be unique and a characteristic subgroup. But then it must also be normal in A_t since A_{t-1} is normal in A_t according to Theorem 20, Chapter 1.

2. **Groups with conformal chains.** We shall say that the chains in a group G are *conformal* when all the complete chains in G have the same length and index type.

We shall first study the solvable groups with conformal chains. These groups may be characterized as follows:

THEOREM 2. *The necessary and sufficient condition for a solvable group to have conformal chains is that the index of consecutive subgroups always be a prime.*

Proof. The definition of a solvable group shows that there exists a chain in which all indices are primes; hence all complete chains must have this property if they are to be conformal.

A consequence of this theorem is

THEOREM 3. *The complete chains in a nilpotent group are conformal.*

Another consequence of Theorem 1 is

THEOREM 4. *In a solvable group the condition that all complete chains shall have the same length implies that they are conformal.*

Proof. The length of the chains must be equal to the number of prime factors in the group order, since there always exists one chain with this property in a solvable group.

We shall now proceed to the deduction of the following important fact:

THEOREM 5. *Every group with conformal chains is solvable.*

Proof. Let us suppose that the theorem were not true. Then there would exist simple groups with conformal chains. Among these we choose one of minimal order. Such a simple group G would have the property that all subgroups were solvable groups with conformal chains.

Now let M be a maximal group in G. We shall show that the index $m = [G : M]$ is a prime number. Let us suppose, namely, that m is composite and that m is divisible by p^α, $\alpha \geqq 1$, while the order of M is divisible by the prime p to the power p^μ. A Sylow group S_p of G corresponding to p has complete chains of length $\mu + \alpha$ and all indices equal to p.

We construct two complete chains in G. The first begins with the $\mu + \alpha$ terms in S_p. The other is obtained by completing a chain in M to a chain in G by adding the group G with the index m. Both chains shall have the same indices in some order. But since the first chain contains $\mu + \alpha$ indices p, while the second only contains μ of them, one must have $\alpha = 1$ and $m = p$.

This shows that all maximal groups in our simple group G must have prime indices. Furthermore, G must have at least two maximal groups with different prime indices. This follows from the

LEMMA. *Any group contains maximal subgroups whose indices are not divisible by any prescribed prime.*

Proof. Any maximal group M containing a Sylow group corresponding to p has an index not divisible by p.

The contradiction proving Theorem 5 is now a direct consequence of the following theorem:

THEOREM 6. *Let G be a finite group containing a maximal group with prime index p. If p is not the greatest prime dividing the group order, then G is composite.*

Proof. G can be represented as a substitution group of degree p. This representation cannot be a true representation since any element whose order is a prime greater than p must correspond to the unit permutation.

It should be observed that when p is the greatest prime dividing the group order, the group may be simple. An example is the alternating group on 5 elements of order 60 which has a subgroup of order 12.

We shall now mention a few other properties of groups with conformal chains. When Theorem 1 is combined with Theorem 13, Chapter IV, one obtains

THEOREM 7. *In a group with conformal chains any arrangement of the prime factors of the group order represents the indices of some chain.*

This implies further

THEOREM 8. *A group with conformal chains has subgroups of every order dividing the group order.*

It is also seen that every subgroup and every quotient group has the same property.

By combining Theorem 7 with Theorem 1, one also sees:

THEOREM 9. *A group with conformal chains is dispersible.*

3. **Groups in which there exists a principal complete chain.** We shall now prove a theorem which brings the groups with conformal chains in connection with another important type of group.

THEOREM 10. *The necessary and sufficient condition that a group G shall have conformal chains is that G contain a principal chain which is complete; i.e., the indices in a principal chain shall be primes.*

Proof. We suppose first that such a principal chain exists. Then G is solvable and from Theorem 4, Chapter III it follows that all indices in any complete chain must be primes. Theorem 2 shows then that G must have conformal chains.

To prove the converse let G be a group with conformal chains. To prove that the indices in a principal chain are primes it is sufficient to show that G contains one normal subgroup of prime order, because the same argument may then be applied to the quotient groups.

According to Theorem 9, G contains a normal Sylow group S_1 of order $p_1^{\alpha_1}$, where p_1 is the largest prime dividing the group order. The p_1-group S_1 has a center C_1 which is a characteristic subgroup of S_1, and hence normal in G.

This shows finally that G contains a minimal normal Abelian subgroup L_1 of order p_1^β and type $(p_1, p_1, \cdots)$ contained in the center of S_1.

Now let K denote a subgroup of G of index $p_1^{\alpha_1}$ and order $N_1 = p_2^{\alpha_2} \cdots p_r^{\alpha_r}$. All such subgroups K are conjugate and for any of them one must have

$$G = K \cup S_1, \qquad K \cap S_1 = E. \tag{7}$$

We consider the subgroup

$$G_1 = K \cup L_1, \qquad K \cap L_1 = E \tag{8}$$

of order $p_1^\beta \cdot N_1$. Since G_1 also has conformal chains, it must have subgroups of every order according to Theorem 8. Let H denote a subgroup of G_1 of order $p_1 \cdot N_1$. According to a previous remark we can always choose a suitable conjugate of H such that $H \supset K$. By the Dedekind relation it follows from (8) that $H = K \cup (H \cap L_1)$. Here the groups K and $H \cap L_1$ are permutable and the order of $H \cap L_1$ is p_1. We shall show that $H \cap L_1$ must be normal in G. First, this group is normal in H and hence $H \cap L_1$ is left invariant by transformation with any element of K. Secondly, $H \cap L_1$ belongs to the center of S_1 and from (7) it follows that $H \cap L_1$ is normal in G.

The proof of Theorem 10 shows incidentally that a group with conformal chains must have a principal chain

$$E \subset N_1 \subset N_2 \subset \cdots \subset N_k = G, \tag{9}$$

where the indices are primes in non-increasing order.

4. **Construction of groups with conformal chains.** Now let G be a group with conformal chains and (9) some complete principal chain. The index $\{N_{i+1} : N_i\} = p$ is a prime and the group of automorphisms is cyclic of order $p - 1$. Any inner automorphism of G defined by some element g will induce some automorphism in N_{i+1}/N_i. If this automorphism is the unit automorphism for all i and for all elements g, then N_{i+1}/N_i belongs to the center of G/N_i and G must be nilpotent according to a well-known criterion for nilpotent groups.[19] This gives for instance

THEOREM 11. *Let G be a group with conformal chains. If the congruence $p \equiv 1 \pmod{q}$ is not satisfied for any pair of primes p and q dividing the group order, then G is nilpotent.*

The criterion for nilpotent groups also gives the following important property of groups with conformal chains:[20]

[19] See pp. 166–167 of the reference in footnote 7.

[20] This theorem may also be considered as a consequence of a result of G. Zappa, *Sui gruppi supersolubili*, Rendiconti del Sem. Mat. di Roma, (4), vol. 2(1938), pp. 323–330. In this paper, which appeared after the completion of my paper, it is shown that a group with complete principal chains (supersolvable group) has a nilpotent commutator group. The proof given above is considerably simpler than the proof given by Zappa. (Added in proof.)

THEOREM 12. *If G is a group with conformal chains, then its commutator group is nilpotent.*

Proof. Since the group of automorphisms of N_{i+1}/N_i is Abelian, any commutator in G must induce the identical automorphism in all such quotient groups. Furthermore, a principal chain in the commutator group C is given by

$$E \subseteq C \cap N_1 \subseteq C \cap N_2 \subseteq \cdots \subseteq C \cap N_k = C.$$

But an element c in C must also induce the unit automorphism in $N_{i+1} \cap C/N_i \cap C$, hence C is nilpotent.

Theorem 12 states that every group with conformal chains can be obtained from a nilpotent group C by extension with an Abelian factor group G/C. Not all such groups will have conformal chains or a complete principal chain, but this property gives an interesting idea of the generality of such groups.

5. **Groups with subgroups of every possible order.** We shall now turn to another characterization of our groups with conformal chains. Let us say for short that a group G has *subgroups of every possible order* if it contains a subgroup of order n for every n dividing the group order N. It follows from the characterization of solvable groups by P. Hall that such groups must be solvable. It is also clear that they must have subgroups of index p for every prime p dividing the group order. This implies, as we have seen before, that certain of the indices in a principal chain must be equal to these primes; in particular G must have a normal subgroup of index p_r, where p_r is the smallest prime dividing the group order.

It would be an interesting problem to determine all groups having subgroups of every possible order, but in spite of the great limitations imposed on the group by this condition it seems rather difficult to obtain a simple characterization of such groups. This remains true even if one supposes that all subgroups shall have this property. In this case one can, however, prove

THEOREM 13. *Let G be a group such that G and every subgroup of G have subgroups of every possible order. Then G is dispersible.*

Proof. G must have a subgroup G_1 whose index is p_r and hence G_1 is normal in G. When the same argument is applied to G_1, the theorem becomes a consequence of Theorem 1.

Theorem 13 implies that G has a normal subgroup whose order is p_1. By induction one concludes

THEOREM 14. *Let G be a group such that every subgroup and every quotient group of G have subgroups of every possible order. Then G is a group with conformal chains, and conversely.*

In connection with this theorem it would be of interest to know whether the condition on the quotient groups is necessary. If it is needed, it might be shown through the construction of a suitable example.

The investigations of this last chapter may be recapitulated as follows:

There exists a type of solvable groups which is characterized completely by each of the following properties:

(α) Groups with conformal chains.

(β) Groups in which all consecutive indices are primes.

(γ) Groups with a complete principal chain.

(δ) Groups such that all subgroups and quotient groups have subgroups of every possible order.

Many very interesting other types of groups are included in this general type of groups, and I hope to return to some of these studies upon a later occasion.

YALE UNIVERSITY.

INVARIANTS OF CERTAIN STOCHASTIC TRANSFORMATIONS: THE MATHEMATICAL THEORY OF GAMBLING SYSTEMS

By Paul R. Halmos

Introduction. The "Regellosigkeit" principle of von Mises has been shown to correspond in the mathematical theory of probability to the fact that certain transformations of infinite dimensional Cartesian space into itself are measure preserving. It is the purpose of this paper to investigate the behavior of such transformations on more general spaces. The theorems at the basis of this work are stated in the first section and applied to obtain results concerning the existence and independence of "Kollektivs" in the second and third sections. In §§4, 5, and 6 certain invariants of the transformations considered are obtained. Previous results on these transformations are shown to be special cases of these invariance theorems.

1. Preliminary definitions and theorems.

In this section we shall define the concepts and state the theorems which are the basis of all the work of the later sections.

Definition 1. A collection $\mathfrak{F}_1$ of sets in a space Ω_1 is a *field* if $E_1 \epsilon \mathfrak{F}_1$ and $E_2 \epsilon \mathfrak{F}_1$ implies $E_1 + E_2 \epsilon \mathfrak{F}_1$ and $E_1 - E_1E_2 \epsilon \mathfrak{F}_1$.[1]

Definition 2. A collection $\mathfrak{B}_1$ of sets in a space Ω_1 is a *Borel field* if $\mathfrak{B}_1$ is a field and if $E_j \epsilon \mathfrak{B}_1$ $(j = 1, 2, \cdots)$ implies $\sum_{j=1}^{\infty} E_j \epsilon \mathfrak{B}_1$.

Definition 3. A *probability measure* is an additive, non-negative set function $P_1(E)$ defined on a field $\mathfrak{F}_1$ in a space Ω_1, with $P_1(\Omega_1) = 1$, such that $P_1(\sum_{j=1}^{\infty} E_j) = \sum_{j=1}^{\infty} P_1(E_j)$ whenever $\{E_j\}$ is a sequence of disjunct sets belonging to $\mathfrak{F}_1$ whose sum is also in $\mathfrak{F}_1$.

Definition 4. A space Ω_1 in which a probability measure P_1 has been defined on a Borel field $\mathfrak{B}_1$ is a *probability space*.

Definition 5. A *measurable set* in a probability space Ω_1 is a set E such that $E \epsilon \mathfrak{B}_1$.

Definition 6. Let Ω_1 be a probability space and Ω_1' a space on which there is given a Borel field $\mathfrak{B}_1'$ of measurable sets. Let $\phi(x)$ be a single-valued function whose domain is Ω_1 and whose range is in Ω_1'. $\phi(x)$ is a *measurable function* if the set E of points $x \epsilon \Omega_1$ for which $\phi(x)$ is in $E' \subseteq \Omega_1'$ is measurable whenever E' is.[2] If $\phi(x)$ is real valued, it is measurable if for every real number λ the

Received May 2, 1938.

[1] All the fields used in this paper will also satisfy the condition that if $E \epsilon \mathfrak{F}_1$, then $CE \epsilon \mathfrak{F}_1$, where CE is the complement in Ω_1 of the set E.

[2] The symbol $\{\phi(x) \epsilon E'\}$ will be used to denote E.

set $\{\phi(x) < \lambda\}$ is measurable. The Borel field of sets on the real line is taken, in this paper, as the collection of Borel sets.

DEFINITION 7. If $\phi(x)$ is a real-valued measurable function on Ω_1, the function $F(\lambda)$ of the real variable λ, $F(\lambda) = P_1\{\phi(x) < \lambda\}$, is its *distribution function.*

DEFINITION 8. Two measurable functions ϕ and ϕ' (defined on Ω_1 and Ω_1', respectively) whose ranges lie in the same space Ω_1'' have the *same distribution* if, for every $E'' \epsilon \mathfrak{B}_1''$, $P_1\{\phi \epsilon E''\} = P_1'\{\phi' \epsilon E''\}$. Thus, in particular, two real-valued measurable functions have the same distribution if and only if their distribution functions are identical.

DEFINITION 9. The class of all measurable functions, with ranges in some fixed space, but not necessarily all defined on the same space, such that every two have the same distribution is a *chance variable.* Any member of this class is a *representation* of the chance variable.

So far we have defined a single chance variable ϕ, in isolation from all other chance variables. If ϕ' is another chance variable (with range in the range space of ϕ), our definition enables us to answer the questions "What is the probability that $\phi \epsilon E$?" and "What is the probability that $\phi' \epsilon E'$?", but it does not give an answer to the question "What is the probability that both $\phi \epsilon E$ and $\phi' \epsilon E'$?" Since chance variables usually present themselves not singly but in sets and are connected with each other in rather special ways, we are led to the following considerations.

Associated with every probability space Ω_1 there is another space Ω defined as follows. Let Ω be the space of all infinite sequences $\omega = \{x_1, x_2, \cdots\}$, where $x_j \epsilon \Omega_1$ $(j = 1, 2, \cdots)$. Let $\mathfrak{B}$ be the smallest Borel field which contains every set determined by conditions of the form $x_j \epsilon E_j$, $E_j \epsilon \mathfrak{B}_1$ $(j = 1, \cdots, n)$. Until we define a probability measure on $\mathfrak{B}$, we may not consider Ω as a probability space. We make, however, the following definition.

DEFINITION 10. If a probability measure P is defined on the Borel field $\mathfrak{B}$ in Ω, the probability space Ω is a *stochastic process* associated with Ω_1.[3]

DEFINITION 11. Let $\alpha_1, \cdots, \alpha_n$ be any finite set of subscripts. The set E is a *cylinder set* over $x_{\alpha_1}, \cdots, x_{\alpha_n}$ if, whenever $\omega \epsilon E$, $\omega = \{x_1, x_2, \cdots\}$, then any point ω', obtained from ω by altering the coördinates $x_{\alpha_1}, \cdots, x_{\alpha_n}$ only, is also in E.

The collection of all measurable cylinder sets over $x_{\alpha_1}, \cdots, x_{\alpha_n}$ forms a Borel field $\mathfrak{B}_{\alpha_1,\dots,\alpha_n} \subseteq \mathfrak{B}$.

DEFINITION 12. If two probability measures are defined on the Borel fields $\mathfrak{B}'$ and $\mathfrak{B}''$ respectively, $\mathfrak{B}' \subseteq \mathfrak{B}$, $\mathfrak{B}'' \subseteq \mathfrak{B}$, they are *coherent* if they assign the same values to sets common to $\mathfrak{B}'$ and $\mathfrak{B}''$.

In terms of the preceding three definitions we are now able to formulate a mathematical description of at least one of the ways in which chance variables occur in physical problems.

DEFINITION 13. A sequence, finite or infinite, of chance variables ϕ_n (with

[3] This is not the most general definition of stochastic process, but it is the one that is to be used exclusively in this paper.

ranges all in the same space Ω_1) is *stochastic* if the following conditions are satisfied.

(i) To every set of conditions of the form $\phi_j \in E_j$, $E_j \in \mathfrak{B}_1$ $(j = 1, \cdots, n)$ there corresponds a number $P_n(\phi_1 \in E_1, \cdots, \phi_n \in E_n)$.

(ii) A unique probability measure may be so defined on the Borel field $\mathfrak{B}_n = \mathfrak{B}_{1,\dots,n}$ that the set $\{x_1 \in E_1\} \cdots \{x_n \in E_n\}$ has measure

$$P_n(\phi_1 \in E_1, \cdots, \phi_n \in E_n).$$

(iii) The probability measures on the Borel fields $\mathfrak{B}_j$ $(j = 1, 2, \cdots)$ are coherent, each with the others.

(iv) The function[4] $x_j(\omega)$ is a representation of ϕ_j $(j = 1, 2, \cdots)$.

The following is a fundamental theorem on the representation of stochastic sequences of chance variables. It was proved in the case where Ω_1 is the real line by Kolmogoroff[5] and in the general case by Doob.[6]

THEOREM 1. *Given a stochastic sequence $\{\phi_n\}$ of chance variables, a unique probability measure P may be so defined on $\mathfrak{B}$ that P is coherent with each P_n $(n = 1, 2, \cdots)$.*

Let Ω_1 be a probability space, and for each n let the chance variable ϕ_n, with domain and range on Ω_1, take the value x at the point $x \in \Omega_1$. It is well known[7] that if to every set of conditions of the form $\phi_j \in E_j$, $E_j \in \mathfrak{B}_1$ $(j = 1, \cdots, n)$ we assign the number $P_1(\phi_1 \in E_1) \cdots P_1(\phi_n \in E_n)$ the sequence $\{\phi_k\}$ is stochastic.

THEOREM 2. *Given any probability space Ω_1, a unique probability measure P may be so defined on Ω that a set determined by the conditions $x_j \in E_j$, $E_j \in \mathfrak{B}_1$ $(j = 1, \cdots, n)$ has the measure $P_1(\phi_1 \in E_1) \cdots P_1(\phi_n \in E_n)$.*

DEFINITION 14. A *system* is a sequence $\{f_n\}$ of measurable functions on the stochastic process Ω satisfying the following conditions.

(i) $f_1(\omega) \equiv 0$ or else $f_1(\omega) \equiv 1$.

(ii) $f_n(\omega)$ $(n > 1)$ depends only on $x_1, \cdots, x_{n-1}$.

(iii) $f_n(\omega)$ $(n = 1, 2, \cdots)$ takes only the values 0 and 1.

(iv) $P(\limsup_{n\to\infty} f_n(\omega) = 1) = 1$.[8]

[4] $x_j(\omega)$ is the function which at the point $\omega = \{x_1, x_2, \cdots\}$ takes the value x_j.

[5] X, p. 27. (See bibliography at the end of this paper.)

[6] VI, §1.

[7] Saks, XII, Chapter 3. The property referred to is the conclusion of Theorem 1 for a finite sequence of chance variables.

[8] A gambling system has been so defined by Doob, IV, p. 365. Essentially the same definition was suggested, independently, by Birnbaum and Schreier, I; Wald, XIII; and Huntemann, IX. This definition describes mathematically our intuitive idea of a gambling system. The player will bet on the outcome of the n-th play if $f_n = 1$, and will refrain from betting otherwise. Condition (ii) states that at each stage the player knows the results of the preceding trials only. Condition (iv) is merely a mathematical convenience which ensures that the probability is one that the player bet an infinite number of times.

With every system on Ω we associate a transformation T, defined almost everywhere on Ω and taking Ω into itself, as follows.

DEFINITION 15. Let $a_n(\omega)$ be the lowest integer satisfying the equation $\sum_{j=1}^{a_n} f_j(\omega) = n$. Then, by condition (iv) in the definition of a system, with almost every ω there is associated an infinite sequence $\{a_n\}$ of subscripts. The *system transformation* T is defined by $T(x_1, x_2, \cdots) = \{x_1', x_2', \cdots\} = \{x_{a_1}, x_{a_2}, \cdots\}$.[9]

DEFINITION 16. Let Ω_1 and Ω_1' be probability spaces and let T_1 be a transformation with domain Ω_1 and range in Ω_1'. T_1 is *measure preserving* if for every measurable set $E' \subseteq \Omega_1'$, the set $E = \{T_1(\omega) \in E'\}$ is measurable and $P_1(E) = P_1'(E')$. We shall also write $T^{-1}(E')$ for E.

DEFINITION 17. If the real-valued measurable function $\phi(x)$ defined on the probability space Ω_1 is summable,[10] $\int_{\Omega_1} \phi \, dP_1$ is the *expectation* of the chance variable represented by ϕ. Usually when we work with an integral on the whole space, we shall omit the range of integration in the symbol. We write $E(\phi)$ for the expectation of ϕ.

DEFINITION 18. Let Ω be a stochastic process, n a positive integer, and Λ a measurable set, $\Lambda \subseteq \Omega$. The set function $P(\Lambda E)$, $E \in \mathfrak{B}_{1,\cdots,n}$, is a probability measure on $\mathfrak{B}_{1,\cdots,n}$ that vanishes whenever $P(E) = 0$. Hence, by a well known theorem on absolutely continuous measures,[11] there exists a non-negative summable function on Ω, uniquely determined except for a set of measure zero and depending on the coördinates $x_1, \cdots, x_n$ only, say $P(x_1, \cdots, x_n; \Lambda)$, such that $P(\Lambda E) = \int_E P(x_1, \cdots, x_n; \Lambda) \, dP$ for every set $E \in \mathfrak{B}_{1,\cdots,n}$. $P(x_1, \cdots, x_n; \Lambda)$ is the *conditional probability* of Λ for given $x_1, \cdots, x_n$.[12]

DEFINITION 19. A stochastic process for which $P(x_1, \cdots, x_{n-1}; x_n \in E)$ is almost everywhere independent of $x_1, \cdots, x_{n-1}$ for every $n = 1, 2, \cdots$, and every $E \in \mathfrak{B}_1$ is *independent*. (A stochastic sequence of chance variables will be called independent if the corresponding stochastic process is.)

DEFINITION 20. An independent stochastic process Ω for which the $x_n(\omega)$ all have the same distribution is *stationary*.

It is readily seen that if $\{\phi_n\}$ is a stochastic sequence of independent chance variables, $P(\phi_1 \in E_1, \cdots, \phi_n \in E_n) = P(\phi_1 \in E_1) \cdots P(\phi_n \in E_n)$ (for $n = 1, 2, \cdots$ and $E_j \in \mathfrak{B}_1$ $(j = 1, \cdots, n)$), and conversely.

In terms of our definitions we can now state the following fundamental theorems.

THEOREM 3. *If $Q(E)$ is an additive, non-negative set function defined on a field $\mathfrak{F}_1$ in a space Ω_1, with $Q(\Omega_1) = 1$, and if $Q(\sum_{j=1}^{\infty} E_j) = \sum_{j=1}^{\infty} Q(E_j)$ whenever*

[9] Theorem 1 shows that the transformed space Ω' is a stochastic process.

[10] For integration in abstract spaces see Saks, XII, Chapter 1.

[11] See Saks, XII, p. 36.

[12] Kolmogoroff, X, Chapter 5.

$\{E_n\}$ is a sequence of disjunct sets belonging to $\mathfrak{F}_1$ whose sum is also in $\mathfrak{F}_1$, then $P(E)$ may be so defined on the smallest Borel field $\mathfrak{B}_1$ including $\mathfrak{F}_1$ that it becomes a probability measure that coincides with Q on sets of $\mathfrak{F}_1$.[13]

THEOREM 4. *Let Ω be an independent, stationary, stochastic process corresponding to a probability space Ω_1 and let E_1 be any measurable set, $E_1 \subseteq \Omega_1$. Then if $g(x)$ is the characteristic function of E_1 on Ω_1, we have*

$$\lim_{n\to\infty} \frac{1}{n} \sum_{j=1}^{n} g(x_j) = P_1(E_1)$$

almost everywhere on Ω.[14]

THEOREM 5. *On an independent, stationary, stochastic process every system transformation is measure preserving.*[15]

2. **Application to "Kollektivs".** Much work has been done in recent years on refining the definition of "Kollektiv" as formulated by von Mises.[16] In this section we shall derive a consequence of Theorem 5 which includes some of the results concerning the existence of certain "admissible numbers" or "Kollektivs".[17]

DEFINITION 21. Let $Q(E)$ be an additive, non-negative set function defined on a field $\mathfrak{F}_1$ in a space Ω_1, with $Q(\Omega_1) = 1$. To every set $E \subseteq \Omega_1$ make correspond the numbers

$$Q^*(E) = \underset{\substack{E_1 \supseteq E \\ E_1 \in \mathfrak{F}_1}}{\text{g.l.b.}}\, Q(E_1) \quad \text{and} \quad Q_*(E) = \underset{\substack{E_1 \subseteq E \\ E_1 \in \mathfrak{F}_1}}{\text{l.u.b.}}\, Q(E_1).$$

Let $\mathfrak{J}_1$ be the collection of sets such that $Q^*(E) = Q_*(E)$. It is readily verified that $\mathfrak{J}_1$ is a field, $\mathfrak{J}_1 \supseteq \mathfrak{F}_1$, and we define for every $E \in \mathfrak{J}_1$, $Q(E) = Q^*(E) = Q_*(E)$. $\mathfrak{J}_1$ is the collection of *Jordan measurable sets* with respect to $\mathfrak{F}_1$.

THEOREM 6. *Let P_1 be a probability measure defined on a denumerable field $\mathfrak{F}_1$ in a space Ω_1. Let $\mathfrak{J}_1$ be the collection of Jordan measurable sets with respect to $\mathfrak{F}_1$, and $\mathfrak{B}_1$ the Borel extension of $\mathfrak{J}_1$. Since P_1 may be defined on $\mathfrak{J}_1$ and then on $\mathfrak{B}_1$ coherently with its definition on $\mathfrak{F}_1$ so that it is a probability measure, Ω_1 becomes a probability space. Let Ω be the independent, stationary, stochastic process associated with Ω_1 and let $T_1, T_2, \cdots$ be any sequence of system transformations on Ω. Finally, write $T_n(x_1, x_2, \cdots) = \{x_1^n, x_2^n, \cdots\}$. Then there is a set Z of measure zero on Ω such that*

$$\lim_{N\to\infty} \frac{1}{N} \sum_{j=1}^{N} g_E(x_j^n) = P_1(E)$$

[13] See, for example, Hahn, VII, p. 433.

[14] This theorem is known as the strong law of large numbers. For a proof see VIII, p. 37, or V, p. 764.

[15] Doob (IV, p. 365) proved this for the case of real-valued chance variables. The proof is the same as for this case. This theorem (which we shall obtain later as a special case of Theorem 10 of this paper) is what corresponds in the classical theory to the von Mises "Regellosigkeit" principle. See XI, p. 14.

[16] See von Mises, XI.

[17] See, for example, Copeland, II and III, and Wald, XIII.

for every $\omega \in \Omega - Z$, *every* $E \in \mathfrak{Z}_1$ ($gE(x)$ *denoting the characteristic function of* E), *and all* $n = 1, 2, \cdots$.

Proof. Let $E_1, E_2, \cdots$ be the total collection of sets in $\mathfrak{F}_1$. By Theorem 4

$$\lim_{N\to\infty} \sum_{j=1}^{N} g_{E_i}(x_j) = P_1(E_i)$$

except on a set A_i of measure zero ($i = 1, 2, \cdots$). Write $A = A_1 + A_2 + \cdots$, and $C_n = T_n^{-1}(A)$. Since A is of measure zero and each T_n is measure preserving (Theorem 5), C_n is also of measure zero. Finally T_n may fail to be defined on a set D_n of measure zero. Write $Z = A + (C_1 + C_2 + \cdots) + (D_1 + D_2 + \cdots)$; then $P(Z) = 0$. Now take ω to be any point $\in \Omega - Z$, E to be any set in $\mathfrak{Z}_1$ and n to be any positive integer. There exist two sequences of sets O_k and I_k such that $I_k \subseteq E \subseteq O_k$, $I_k \in \mathfrak{F}_1$, $O_k \in \mathfrak{F}_1$ ($k = 1, 2, \cdots$), and $\lim_{k\to\infty} P_1(I_k) = \lim_{k\to\infty} P_1(O_k) = P_1(E)$. The sets I_k and O_k are among the sets $E_1, E_2, \cdots$, hence we have for all k

$$\lim_{N\to\infty} \frac{1}{N} \sum_{j=1}^{N} g_{I_k}(x_j) = P_1(I_k) \quad \text{and} \quad \lim_{N\to\infty} \frac{1}{N} \sum_{j=1}^{N} g_{O_k}(x_j) = P_1(O_k),$$

since ω was taken out of $Z \supseteq A_1 + A_2 + \cdots$. Also $T_n(\omega)$ is not in $A_1 + A_2 + \cdots$, for otherwise ω would be contained in some C_n, and this contradicts the assumption that $\omega \in \Omega - Z$. Hence the last written relations hold when x_j is replaced by x_j^n; thus

$$\lim_{N\to\infty} \frac{1}{N} \sum_{j=1}^{N} g_{I_k}(x_j^n) = P_1(I_k) \quad \text{and} \quad \lim_{N\to\infty} \frac{1}{N} \sum_{j=1}^{N} g_{O_k}(x_j^n) = P_1(O_k).$$

Since $I_k \subseteq E \subseteq O_k$, we have $g_{I_k}(x) \leqq g_E(x) \leqq g_{O_k}(x)$ for all $x \in \Omega_1$. Hence

$$\frac{1}{N} \sum_{j=1}^{N} g_{I_k}(x_j^n) \leqq \frac{1}{N} \sum_{j=1}^{N} g_E(x_j^n) \leqq \frac{1}{N} \sum_{j=1}^{N} g_{O_k}(x_j^n);$$

whence

$$P(I_k) = \lim_{N\to\infty} \frac{1}{N} \sum_{j=1}^{N} g_{I_k}(x_j^n) \leqq \liminf_{N\to\infty} \frac{1}{N} \sum_{j=1}^{N} g_E(x_j^n)$$
$$\leqq \limsup_{N\to\infty} \frac{1}{N} \sum_{j=1}^{N} g_E(x_j^n) \leqq \lim_{N\to\infty} \frac{1}{N} \sum_{j=1}^{N} g_{O_k}(x_j^n) = P_1(O_k).$$

Since this is true for all k, we have

$$P_1(E) \leqq \liminf \leqq \limsup \leqq P_1(E),$$

so that

$$\liminf = \limsup = \lim_{N\to\infty} \frac{1}{N} \sum_{j=1}^{N} g_E(x_j^n) = P_1(E).^{18}$$

[18] If we exclude trivial cases by insisting that P_1 have at least two different positive values, the set of points ω for which the conclusion of the theorem holds always has the power of the continuum.

3. **The von Mises definition of independence.** The axioms and definitions of the theory of probability as expounded by von Mises have been shown to correspond to theorems in the classical theory. Thus, for example, the various formulations of the law of large numbers correspond to the frequency definition of probability, and Theorem 5 of this paper corresponds to the "Regellosigkeit" principle. We shall now prove a theorem that expresses the fundamental idea in the von Mises definition of independence. There is a close analogy between this theorem and Theorem 5, and the method of proof here employed is similar to that of Doob.[19]

Let Ω be any stochastic process and Ω' an independent and stationary stochastic process. Let Ω^2 be the space of all sequences $\omega^2 = \{(x_1, y_1), (x_2, y_2), \cdots\}$, where $\{y_1, y_2, \cdots\} \epsilon \Omega$ and $\{x_1, x_2, \cdots\} \epsilon \Omega'$. A unique probability measure is defined on Ω^2 by the conditions

$$P\{(x_1 \epsilon E_1) \cdots (x_n \epsilon E_n)(y_1 \epsilon F_1) \cdots (y_n \epsilon F_n)\}$$
$$= P(x_1 \epsilon E_1) \cdots P(x_n \epsilon E_n)P\{(y_1 \epsilon F_1) \cdots (y_n \epsilon F_n)\}.$$

THEOREM 7. *Let $E_1, E_2, \cdots$ be an infinite sequence of measurable sets in the range space of Ω such that the probability is one that an infinite number of the conditions $(y_n \epsilon E_n)$ are satisfied. Let $a_n(\omega^2)$ be the n-th subscript such that $y_{a_n} = E_{a_n}$. To every point ω^2 of Ω^2 make correspond the point $\omega' \epsilon \Omega'$, $\omega' = \{x_1', x_2', \cdots\}$, where $x_n' = x_{a_n}$. The transformation T so defined is defined almost everywhere on Ω^2 taking values in Ω' and is measure preserving in the sense that if Λ' is any measurable set in Ω' and $\Lambda^2 = T^{-1}(\Lambda')$ is the total set of points of Ω^2 whose images are in Λ', then Λ^2 is measurable and $P(\Lambda^2) = P(\Lambda')$.*

Proof. It is sufficient to prove the theorem for sets Λ' of the form

$$\Lambda' = \{x_1' \epsilon E_1', \cdots, x_n' \epsilon E_n'\}, \tag{1}$$

where $E_1', \cdots, E_n'$ are measurable sets in the range space of Ω'. We proceed by induction.

Let $n = 1$; $\Lambda' = \{x_1' \epsilon E_1'\}$. We have, excepting always the set of measure zero where T may not be defined,

$$\Lambda^2 = \{(y_1 \epsilon E_1)(x_1 \epsilon E_1')\} + \{(y_1 \epsilon CE_1)(y_2 \epsilon E_2)(x_2 \epsilon E_1')\} + \cdots. \tag{2}$$

Hence, since the sets in this sum are disjunct we have, by the definition of measure on Ω^2,

$$P(\Lambda^2) = P(x_1 \epsilon E_1')P(y_1 \epsilon E_1) + P(x_2 \epsilon E_1')P\{(y_1 \epsilon CE_1)(y_2 \epsilon E_2)\} + \cdots. \tag{3}$$

The last written expression is the product of $P(x_n \epsilon E_1')$ by the probability that at least one of the conditions $(y_n \epsilon E_n)$ is satisfied, hence $P(\Lambda^2) = P(x_n \epsilon E_1') = P(\Lambda')$.

This proves the theorem for $n = 1$. Assume that it is true for $n - 1$. Write

[19] IV.

$\Lambda' = \{x_1' \epsilon E_1', \cdots, x_n' \epsilon E_n'\}$, $\Lambda^2 = \{T^{-1}(\Lambda')\}$. In order to prove that Λ^2 is measurable consider the functions $x_j'(\omega^2) = x_{a_j}(\omega^2)$.

$$\{x_{a_j}(\omega^2) \epsilon E\} = \{a_j = j\}\{x_j \epsilon E\} + \{a_j = j + 1\}\{x_{j+1} \epsilon E\} + \cdots .$$

The sets $\{x_m \epsilon E\}$ are measurable, if E is any measurable set in the range space of Ω'. The set $\{a_j = m\}$ is the set where the j-th y-coördinate of ω^2 which belongs to its E_i is y_m; this set is the sum of sets where y_m is in E_m and of the preceding y's precisely $j - 1$ are in their E_i. These summands, and hence their sum, and therefore, finally, the set $\{x_j' \epsilon E\}$ are measurable. Hence Λ, being the intersection of the n sets $\{x_j' \epsilon E_j'\}$, is measurable.

Denote by Λ_0 the set $\{x_j' \epsilon E_j' \ (j = 1, \cdots, n-1)\}$. Then

$$\Lambda^2 = \sum_{j=n}^{\infty} \{\omega^2 \epsilon \Lambda_0, a_n = j, x_j \epsilon E_n'\}. \tag{4}$$

Consider any summand in (4): $\{\omega^2 \epsilon \Lambda_0, a_n = j, x_j \epsilon E_n'\}$. It is readily verified that the set $\{\omega^2 \epsilon \Lambda_0, a_n = j\}$ is a cylinder set over the first $j - 1$ (x, y)-coördinates of ω^2. Hence

$$P(\omega^2 \epsilon \Lambda_0, a_n = j, x_j \epsilon E_n') = P(\omega^2 \epsilon \Lambda_0, a_n = j)P(x_j \epsilon E_n'). \tag{5}$$

From (4) and (5), the stationary character of Ω', and the disjunct nature of the summands in (4) we obtain

$$P(\Lambda) = P(x_n \epsilon E_n') \sum_{j=n}^{\infty} P(\omega^2 \epsilon \Lambda_0, a_n = j). \tag{6}$$

The last written summation is the measure of the intersection of Λ_0 with the set where at least one of the conditions $(y_n \epsilon E_n)$ is satisfied for $n > j$: the measure of the latter set is one; hence

$$P(\Lambda) = P(x_n \epsilon E_n') \cdot P(\Lambda_0). \tag{7}$$

The theorem follows immediately from the induction hypothesis.

4. **The invariance of expectation.** Theorem 5 asserts that under certain hypotheses all probability relations are invariant under a system transformation. In this section we make less restrictive hypotheses on the stochastic process to obtain the conclusion (Theorem 8) that the "fairness" of a gambling game of which the stochastic process is a mathematical description is invariant under a system—where the criterion for fairness is expressed, as usual, in terms of the vanishing of certain expectations.

Definition 22. Let Ω be a stochastic process associated with a probability space consisting of real numbers; let n be a positive integer; and suppose that $x_j(\omega)$ is summable for $j = 1, 2, \cdots$. $Q(E) = \int_E x_n(\omega)\, dP$, $E \epsilon \mathfrak{B}_{1,\cdots,(n-1)}$, is a finite, completely additive set function on $\mathfrak{B}_{1,\cdots,(n-1)}$ that vanishes whenever $P(E) = 0$. Hence there exists a summable function on Ω, uniquely deter-

mined except for a set of measure zero and depending on the coördinates $x_1, \cdots, x_{n-1}$ only, say $E(x_1, \cdots, x_{n-1}; x_n)$, such that

$$\int_E x_n(\omega)\,dP = \int_E E(x_1, \cdots, x_{n-1}; x_n)\,dP$$

for every set $E \in \mathfrak{B}_{1,\cdots,(n-1)}$. $E(x_1, \cdots, x_{n-1}; x_n)$ is the *conditional expectation of* x_n for given $x_1, \cdots, x_{n-1}$.

THEOREM 8. *Let Ω be a real-valued stochastic process in which the functions $x_n(\omega)$ are uniformly bounded. Let T be a system transformation on Ω taking $\omega = \{x_1, x_2, \cdots\}$ into $\omega' = \{x_1', x_2', \cdots\}$. If $E(x_1, \cdots, x_{n-1}; x_n)$ vanishes for all n and almost all $x_1, \cdots, x_{n-1}$, then $E(x_1', \cdots, x_{n-1}'; x_n')$ vanishes for all n and almost all $x_1', \cdots, x_{n-1}'$.*

Proof. Our hypothesis is that

$$\int_M E(x_1, \cdots, x_{n-1}; x_n)\,dP = \int_M x_n\,dP = 0 \tag{8}$$

for all n and all measurable cylinder sets M over $x_1, \cdots, x_{n-1}$. We are to prove that

$$\int_{M'} E(x_1', \cdots, x_{n-1}'; x_n')\,dP = \int_{M'} x_n'\,dP = 0 \tag{9}$$

for all n and all measurable cylinder sets M' over $x_1', \cdots, x_{n-1}'$. It is sufficient to prove this result for sets M' of the form

$$M' = \{x_1' \in E_1, \cdots, x_{n-1}' \in E_{n-1}\},$$

where $E_1, \cdots, E_{n-1}$ are Borel sets on the real line.[20]

We start then with M' and derive an expression for $P(M'\{x_n' \in E_n\})$, where E_n is a Borel set on the real line. We have

$$\begin{aligned} \{x_m' \in E\} &= \{a_m = m\}\{x_m \in E\} + \{a_m = m+1\}\{x_{m+1} \in E\} + \cdots \\ &= \sum_{j=0}^{\infty} \{a_m = m+j\}\{x_{m+j} \in E\}. \end{aligned} \tag{10}$$

Hence

$$\begin{aligned} M'\{x_n' \in E_n\} &= \{x_1' \in E_1\}\{x_2' \in E_2\} \cdots \{x_{n-1}' \in E_{n-1}\}\{x_n' \in E_n\} \\ &= \Big(\sum_{j_1=0}^{\infty} \{a_1 = j_1 + 1\}\{x_{j_1+1} \in E_1\}\Big) \cdots \Big(\sum_{j_n=0}^{\infty} \{a_n = j_n + n\}\{x_{j_n+n} \in E_n\}\Big) \\ &= \sum_{j_n=0}^{\infty} \cdots \sum_{j_1=0}^{\infty} \{a_1 = j_1 + 1\} \cdots \{a_n = j_n + n\}\{x_{j_1+1} \in E_1\} \\ &\qquad \cdots \{x_{j_n+n} \in E_n\}.^{21} \end{aligned} \tag{11}$$

[20] Every measurable set over $x_1', \cdots, x_{n-1}'$ can be obtained from sets such as M' by at most a denumerable sequence of sum, product, and complement processes. If the integral vanishes for all sets such as M', it will vanish for sums of such sets, etc.

[21] $a_1(\omega), a_2(\omega), \cdots$ is the subscript sequence associated with T, as in Definition 15.

The last expression follows from the ordinary algebra of point sets. We make certain easily verifiable remarks about the last written sum.

(12) The sets being added are disjunct in pairs.

(13) Since, for example, a_2 is always *greater than* a_1, the set $\{a_1 = j_1 + 1\}$ $\{a_2 = j_2 + 2\}$ is empty if $j_1 + 1 \geqq j_2 + 2$; i.e., it is empty when $j_1 > j_2$.

In virtue of (13) we have from (11)

$$M'\{x'_n \in E_n\} = \sum_{j_n=0}^{\infty} \sum_{j_{n-1}=0}^{j_n} \cdots \sum_{j_1=0}^{j_2} \{a_1 = j_1 + 1\} \cdots \{a_n = j_n + n\}\{x_{j_1+1} \in E_1\} \cdots \{x_{j_n+n} \in E_n\}. \tag{14}$$

From this we have, in virtue of (12),

$$P(M'\{x'_n \in E_n\}) = \sum_{j_n=0}^{\infty} \cdots \sum_{j_1=0}^{j_2} P(\{a_1 = j_1 + 1\} \cdots \{a_n = j_n + n\}\{x_{j_1+1} \in E_1\} \cdots \{x_{j_n+n} \in E_n\}). \tag{15}$$

Let us write

$$M_{j_1 \cdots j_n} = \{a_1 = j_1 + 1\} \cdots \{a_n = j_n + n\}\{x_{j_1+1} \in E_1\}\{x_{j_{n-1}+n-1} \in E_{n-1}\}.$$

It is immediately verifiable that the set $M_{j_1 \cdots j_n}$ is a cylinder set over the coördinates $x_1, \cdots, x_{j_n+n-1}$.[22] Write also

$$M_{j_n} = \sum_{j_{n-1}=0}^{j_n} \cdots \sum_{j_1=0}^{j_2} M_{j_1 \cdots j_n}.$$

Since we already saw that the summands are disjunct sets, we have

$$P(M_{j_n}) = \sum_{j_{n-1}=0}^{j_n} \cdots \sum_{j_1=0}^{j_2} P(M_{j_1 \cdots j_n}). \tag{16}$$

Since the x_n were assumed to be uniformly bounded, the x'_n will be uniformly bounded, and therefore summable.

$$\int_{M'} x'_n \, dP = \lim_{\delta \to 0} \sum_{r=-\infty}^{\infty} r\delta P(M'\{r\delta \leqq x'_n < (r+1)\delta\}). \tag{17}$$

Then, from (15)

$$\int_{M'} x'_n \, dP = \lim_{\delta \to 0} \sum_{r=-\infty}^{\infty} r\delta \sum_{j_n=0}^{\infty} P(M_{j_n}\{r\delta \leqq x_{j_n+n} < (r+1)\delta\}). \tag{18}$$

We may now define the function $y(\omega)$ to be equal to x_{j_n+n} on M_{j_n} ($j_n = 0, 1, 2, \cdots$), and to be zero elsewhere. (The sets M_{j_n} are disjunct sets. The function $y(\omega)$ really depends on n, but since n is being held fixed in this discussion we do not indicate this dependence.) Then

[22] See (ii), Definition 14.

$$
\begin{aligned}
\int_{M'} x'_n\,dP &= \lim_{\delta\to 0} \sum_{r=-\infty}^{\infty} r\delta \sum_{j_n=0}^{\infty} P(M_{j_n}\{r\delta \leqq y < (r+1)\delta\}) \\
&= \lim_{\delta\to 0} \sum_{r=-\infty}^{\infty} r\delta P\,\Big(\sum_{j_n=0}^{\infty} M_{j_n}\{r\delta \leqq y < (r+1)\delta\}\Big) \\
&= \int_{\sum_{j_n=0}^{\infty} M_{j_n}} y\,dP = \sum_{j_n=0}^{\infty} \int_{M_{j_n}} y\,dP = \sum_{j_n=0}^{\infty} \int_{M_{j_n}} x_{j_n+n}\,dP \\
&= \sum_{j_n=0}^{\infty} \lim_{\delta\to 0} \sum_{r=-\infty}^{\infty} r\delta P(M_{j_n}\{r\delta \leqq x_{j_n+n} < (r+1)\delta\}) \qquad (19) \\
&= \sum_{j_n=0}^{\infty} \lim_{\delta\to 0} \sum_{r=-\infty}^{\infty} r\delta \sum_{j_{n-1}=0}^{j_n} \cdots \sum_{j_1=0}^{j_2} P(M_{j_1\cdots j_n}\{r\delta \leqq x_{j_n+n} < (r+1)\delta\}) \\
&= \sum_{j_n=0}^{\infty} \sum_{j_{n-1}=0}^{j_n} \cdots \sum_{j_1=0}^{j_2} \lim_{\delta\to 0} \sum_{r=-\infty}^{\infty} r\delta P(M_{j_1\cdots j_n}\{r\delta \leqq x_{j_n+n} < (r+1)\delta\}) \\
&= \sum_{j_n=0}^{\infty} \sum_{j_{n-1}=0}^{j_n} \cdots \sum_{j_1=0}^{j_2} \int_{M_{j_1}} x_{j_n+n}\,dP = 0.
\end{aligned}
$$

This concludes the proof of Theorem 8.

We remark on the analogy between this theorem and Theorem 5. Here we proved, essentially, that if $E(x_1, \cdots, x_{n-1}; x_n)$ has a constant value independent of n and $x_1, \cdots, x_{n-1}$, then $E(x'_1, \cdots, x'_{n-1}; x'_n)$ will also have that constant value independent of n and $x'_1, \cdots, x'_{n-1}$. Theorem 5, on the other hand, may be phrased as follows. If $P(x_1, \cdots, x_{n-1}; x_n \epsilon E)$ has a constant value depending only on E, but not on n or $x_1, \cdots, x_{n-1}$, then $P(x'_1, \cdots, x'_{n-1}; x'_{n-1} \epsilon E)$ will also have that constant value dependent only on E, but not on n or $x'_1, \cdots, x'_{n-1}$.

5. **The invariance of asymptotic independence.**

Definition 23. The stochastic process Ω is *uniformly asymptotically independent* if there exists a probability measure F on its range space such that $P(x_1, \cdots, x_{n-1}; x_n \epsilon E)$ converges uniformly in ω (but not necessarily in E) almost everywhere to $F(E)$, where E is any measurable set.

The first purpose of this section is to prove the following theorem.

Theorem 9. *The property of uniform asymptotic independence is invariant under every system transformation.*

Instead of proving this theorem we shall state and prove the following slightly stronger one.

Theorem 10. *If for some $\epsilon > 0$, measurable set E, and positive integer n_0, $|P(x_1, \cdots, x_{n-1}; x_n \epsilon E) - F(E)| < \epsilon$ for all $n \geqq n_0$ almost everywhere, then $|P(x'_1, \cdots, x'_{n-1}; x'_n \epsilon E) - F(E)| < \epsilon$ for all $n \geqq n_0$ almost everywhere (where the x'_n are obtained from the x_n by a system transformation).*

The important difference between this theorem and the preceding one is that here we definitely state that the ϵ and n_0 after the system transformation are the same as before.

Proof. Let I'_{n-1} be any measurable cylinder set over $x'_1, \cdots, x'_{n-1}$. Then

$$\left| \int_{I'_{n-1}} P(x'_1, \cdots, x'_{n-1}; x'_n \epsilon E) - F(E)\, dP \right| = | P(I'_{n-1}\{x'_n \epsilon E\}) - P(I'_{n-1})F(E)|$$

$$= \left| \sum_{j=0}^{\infty} P(I'_{n-1}\{a_n = n + j\}\{x_{n+j} \epsilon E\}) - F(E)P(I'_{n-1}) \right|$$

$$\leqq \sum_{j=0}^{\infty} \int_{I'_{n-1}\{a_n = n+j\}} | P(x_1, \cdots, x_{n+j-1}; x_{n+j} \epsilon E) - F(E)|\, dP \tag{20}$$

$$\leqq \sum_{j=0}^{\infty} \int_{I'_{n-1}\{a_n = n+j\}} \epsilon\, dP = \int_{I'_{n-1}} \epsilon\, dP.$$

Since this is true uniformly in I'_{n-1}, we have $| P(x'_1, \cdots, x'_{n-1}; x'_n \epsilon E) - F(E) | < \epsilon$ for $n \geqq n_0$ almost everywhere. This concludes the proof of Theorem 10, and therefore of Theorem 9.

The reason for stating Theorem 10 in its present form is that in this form Theorem 5 is easily seen to follow from it. For, according to the hypotheses of Theorem 5, the hypotheses of Theorem 10 are satisfied with $n_0 = 1$ and every $\epsilon > 0$. The conclusion of Theorem 10 then assures us that after any system transformation the conditions are still satisfied, with $n_0 = 1$ and every $\epsilon > 0$. This implies that the transformed process is independent, stationary, and has the same distributions as the original stochastic process.

Besides the situation just mentioned, there are many other important examples of uniformly asymptotically independent stochastic processes. It is clear, however, that uniform asymptotic independence is a very strong condition. There may be stochastic processes which from a practical point of view are independent in the long run, without being uniformly asymptotically independent in the sense of Definition 23. In order to investigate such processes we make the following definitions.

Definition 24. The stochastic process Ω is *asymptotically independent in probability* if there exists a probability measure F on its range space such that $P(x_1, \cdots, x_{n-1}; x_n \epsilon E)$ converges in probability to $F(E)$, where E is any measurable set. (That is: to every positive number ϵ and measurable set E there corresponds a positive integer n_0 such that $P\{| P(x_1, \cdots, x_{n-1}; x_n \epsilon E) - F(E) | > \epsilon\} < \epsilon$ for $n > n_0$.)

Definition 25. The stochastic process Ω is *ϵ-multiplicative* if there exists a probability measure F on its range space such that to every positive number ϵ there corresponds a positive integer n_0 such that

$$| P(\Lambda\{x_n \epsilon E\}) - P(\Lambda)F(E) | < \epsilon \qquad \text{for } n > n_0,$$

where E is a preassigned measurable set, uniformly in the measurable cylinder set Λ over $x_1, \cdots, x_{n-1}$.

Before investigating the invariance under system transformations of asymptotic independence in probability, we need the following auxiliary theorem.

THEOREM 11. *A necessary and sufficient condition that a stochastic process be asymptotically independent in probability is that it be ϵ-multiplicative.*

Proof. We use the notation of Definitions 24 and 25. We first prove that if the process is ϵ-multiplicative, then it is asymptotically independent in probability. Suppose that the conclusion of the theorem is false: then there exists a positive number ϵ such that

$$P\{| P(x_1, \cdots, x_{n-1}; x_n \epsilon E) - F(E) | > \epsilon\} \geqq \epsilon \tag{21}$$

for an infinite number of values of n. This implies that either there is an infinite number of values of n for which the difference is $> \epsilon$ and positive on a set of measure $\geqq \epsilon$, or else that there is an infinite number of values of n for which the difference is negative and $< -\epsilon$ on a set of measure $\geqq \epsilon$. In either case there is an infinite number of values of n corresponding to which we may find a measurable cylinder set Λ_n over $x_1, \cdots, x_{n-1}$ such that $P(\Lambda_n) \geqq \epsilon$ and such that the difference in (21) is of one sign on all Λ_n and is greater in absolute value than ϵ. For all these values of n we have

$$\begin{aligned} | P(\Lambda_n\{x_n \epsilon E\}) - P(\Lambda_n)F(E) | &= \left| \int_{\Lambda_n} P(x_1, \cdots, x_{n-1}; x_n \epsilon E) - F(E)\, dP \right| \\ &= \int_{\Lambda_n} | P(x_1, \cdots, x_{n-1}; x_n \epsilon E) - F(E) |\, dP \geqq \epsilon P(\Lambda_n) \geqq \epsilon^2. \end{aligned}$$

Since this contradicts the assumption of ϵ-multiplicativity, the sufficiency of the condition is proved.

Assume now that the process is asymptotically independent in probability. Then

$$\begin{aligned} | P(\Lambda\{x_n \epsilon E\}) - P(\Lambda)F(E) | &= \left| \int_{\Lambda} P(x_1, \cdots, x_{n-1}; x_n \epsilon E) - F(E) dP \right| \\ &\leqq \int_{\Lambda} | P(x_1, \cdots, x_{n-1}; x_n \epsilon E) - F(E) |\, dP \\ &= \int_{\Lambda^\epsilon} + \int_{\Lambda_\epsilon} | \cdots |\, dP, \end{aligned}$$

where Λ_ϵ is the part of Λ on which the integrand is $\leqq \epsilon$, and $\Lambda^\epsilon = \Lambda - \Lambda_\epsilon$. For n sufficiently large we have, using asymptotic independence in probability,

$$| P(\Lambda\{x_n \epsilon E\}) - P(\Lambda)F(E) | \leqq \epsilon P(\Lambda_\epsilon) + 2P(\Lambda^\epsilon) \leqq \epsilon + 2\epsilon$$

(since $| P(x_1, \cdots, x_{n-1}; x_n \epsilon E) - F(E) | \leqq 2$). This concludes the proof of Theorem 11.

We are now able to begin the examination of the behavior of asymptotic independence in probability under system transformations.

THEOREM 12. *Let Ω be a stochastic process and let F be a probability measure on its range space such that $P(x_1, \cdots, x_{n-1}; x_n \epsilon E)$ converges almost everywhere to $F(E)$, where E is any measurable set. Then, if T is any system transformation on Ω, the transformed process $T(\Omega)$ is asymptotically independent in probability.*

We note that the conditions put on Ω and those that $T(\Omega)$ is proved to satisfy are not the same. The process Ω is asymptotically independent in probability, and a little more: the conditional probability distributions are assumed not merely to converge in probability but to converge almost everywhere; but the process $T(\Omega)$ is merely asserted to be asymptotically independent in probability. That this lack of symmetry is in a certain sense in the nature of things will be shown later.

Proof. According to Theorem 11 it is sufficient to prove that $T(\Omega)$ is ϵ-multiplicative; that is, it is sufficient to show that, for every ϵ, $| P(\Lambda'\{x'_n \epsilon E\}) - P(\Lambda')F(E) | < \epsilon$, for n sufficiently large, uniformly in the measurable cylinder set Λ' over $x'_1, \cdots, x'_{n-1}$. We have, using our usual notation for system transformations,

$$P(\Lambda'\{x'_n \epsilon E\}) = \sum_{j=0}^{\infty} P(\Lambda'\{a_n = n + j\}\{x_{n+j} \epsilon E\}). \tag{22}$$

Hence

$$\begin{aligned} &| P(\Lambda'\{x'_n \epsilon E\}) - P(\Lambda')F(E) | \\ &\quad = \left| \sum_{j=0}^{\infty} P(\Lambda'\{a_n = n + j\}\{x_{n+j} \epsilon E\}) - P(\Lambda'\{a_n = n + j\})F(E) \right| \\ &\quad = \left| \sum_{j=0}^{\infty} \int_{\Lambda'\{a_n = n+j\}} P(x_1, \cdots, x_{n+j-1}; x_{n+j} \epsilon E) - F(E)\, dP \right| \\ &\quad \leq \sum_{j=0}^{\infty} \int_{\Lambda'\{a_n = n+j\}} | P(x_1, \cdots, x_{n+j-1}; x_{n+j} \epsilon E) - F(E) |\, dP. \end{aligned} \tag{23}$$

Since $P(x_1, \cdots, x_{n-1}; x_n \epsilon E)$ converges to $F(E)$, it converges, by Egoroff's theorem,[23] uniformly on a set D, with $P(D)$ arbitrarily close to one. Hence for given $\epsilon > 0$ we may select D so that $P(CD) = 1 - P(D) < \epsilon$. On D we have

$$| P(x_1, \cdots, x_{n+j-1}; x_{n+j} \epsilon E) - F(E) | < \epsilon \tag{24}$$

for n sufficiently large and $j = 0, 1, 2, \cdots$. Hence, from (23)

$$\begin{aligned} &| P(\Lambda'\{x'_n \epsilon E\}) - P(\Lambda')F(E) | \\ &\leq \sum_{j=0}^{\infty} \int_{\Lambda'\{a_n = n+j\}D} + \int_{\Lambda'\{a_n = n+j\}CD} | P(x_1, \cdots, x_{n+j-1}; x_{n+j} \epsilon E) - F(E) |\, dP \\ &\leq \sum_{j=0}^{\infty} \int_{\Lambda'\{a_n = n+j\}D} \epsilon\, dP + \sum_{j=0}^{\infty} \int_{\Lambda'\{a_n = n+j\}CD} 2\, dP \\ &= \epsilon P(\Lambda'D) + 2P(\Lambda' \cdot CD) \leq \epsilon + 2\epsilon, \end{aligned} \tag{25}$$

as was to be proved.

[23] Saks, XII, p. 18.

THEOREM 13. *Under the hypotheses of Theorem 12, to every positive integer k, measurable sets $E_1, \cdots, E_k$, and positive number ϵ there corresponds a positive integer n_0 such that*

$$| P(\Lambda'\{x'_{n+1} \epsilon E_1\} \cdots \{x'_{n+k} \epsilon E_k\}) - F(E_1) \cdots F(E_k)P(\Lambda') | < \epsilon$$

for all $n > n_0$, uniformly in the measurable cylinder set Λ' over $x'_1, \cdots, x'_n$.

This slightly more general theorem follows readily from Theorem 12 by mathematical induction.

In order to show that without some extra hypothesis (such as convergence in Theorem 12) we may not assert asymptotic independence in probability about $T(\Omega)$, we construct an example of a stochastic process which is asymptotically independent in probability, but which loses this property under a suitable system transformation.

Take $0 < q < p < 1$. Let Ω be a stochastic process in which each x_n takes only the values 0 and 1, and in which the following conditions are satisfied. For $2^m \leqq n \leqq 2^{m+1} - 1$, $P(x_1, \cdots, x_{n-1}; x_n = 0)$ depends only on the coordinates $x_1, \cdots, x_m$ $(m = 0, 1, 2, \cdots)$. For each positive integer m arrange the 2^m possible sets of values of $(x_1, \cdots, x_m)$ in some order: say, for definiteness, that these sets are ordered according to the magnitude of the dyadic fractions $.x_1x_2 \cdots x_m$. Let $P(x_1 = 0) = p$. Let $P(x_1, \cdots, x_{2^m-1}; x_{2^m} = 0)$ be p when $(x_1, \cdots, x_m)$ is the first set in this ordering and let $P(x_1, \cdots, x_{2^m-1}; x_{2^m} = 0)$ be q otherwise. Let $P(x_1, \cdots, x_{2^m}; x_{2^m+1} = 0)$ be p when $(x_1, \cdots, x_m)$ is the second set and q otherwise; and so on, for each n between 2^m and $2^{m+1} - 1$, and every $m = 0, 1, 2, \cdots$. According to a theorem of Doob's[24] these conditions are consistent. That is, if we define $P(x_1, \cdots, x_{n-1}; x_n = 1)$ to be $1 - P(x_1, \cdots, x_{n-1}; x_n = 0)$, then there is a probability measure on Ω which has precisely these functions for its conditional probability functions.

We have to show that this process is asymptotically independent in probability. Let ϵ be any positive number, $0 < \epsilon < p - q$. The probability measure of the set where $| P(x_1, \cdots, x_{n-1}; x_n = 0) - q | > \epsilon$ is the measure of the set where $P(x_1, \cdots, x_{n-1}; x_n = 0) = p$. For $2^m \leqq n \leqq 2^{m+1} - 1$, this set is of the form $\{x_1 = x_1^0, \cdots, x_m = x_m^0\}$. We are to prove then that the measure of such a set approaches zero with n^{-1}. But

$$P(x_1 = x_1^0, \cdots, x_m = x_m^0) = P(x_1 = x_1^0, \cdots, x_{m-1} = x_{m-1}^0) \cdot P(x_1^0, \cdots, x_{m-1}^0; x_m = x_m^0).$$

The last-written conditional probability is not greater than $r = \max \{p, q, (1 - p), (1 - q)\}$; whence $P(x_1 = x_1^0, \cdots, x_m = x_m^0) \leqq r^m$. Hence $P(x_1, \cdots, x_{n-1}; x_n = 0)$ converges in measure to q.

We now define a system transformation T by defining the sequence of choice functions $\{f_n\}$.[25] Let $f_n(x_1, \cdots, x_{n-1})$ be 1 or 0 according as $P(x_1, \cdots, x_{n-1}; x_n = 0) = p$ or q. It is readily verified that this sequence of functions satisfies all the conditions of Definition 14.

[24] VI, §4.
[25] Definition 14.

For $2^m \leqq n \leqq 2^{m+1} - 1$, f_n depends on the coördinates $x_1, \cdots, x_m$ only ($m = 0, 1, 2, \cdots$). Thus $f_1 \equiv 1$; $f_2 = 1$ if $x_1 = 0$ and $f_2 = 0$ otherwise; $f_3 = 0$ if $x_1 = 0$ and $f_3 = 1$ otherwise. Hence we see that a_2 takes only the values 2 and 3 and takes these values on the sets $\{P(x_1; x_2 = 0) = p\}$ and $\{P(x_1, x_2; x_3 = 0) = p\}$, respectively. It is similarly verified that for all n, $2^{n-1} \leqq a_n \leqq 2^n - 1$, and that the set where a_n takes one of its values, say $\{a_n = k\}$, coincides with the set $\{P(x_1, \cdots, x_{k-1}; x_k = 0) = p\}$.

It is now easy to see that $T(\Omega)$ is not asymptotically independent in probability. For if it were, $P(x_n' = 0)$ would certainly converge to q. But

$$
\begin{aligned}
P(x_n' = 0) - q &= \sum_{j=0}^{\infty} P(\{a_n = n + j\}\{x_{n+j} = 0\}) - P(a_n = n + j)q \\
&= \sum_{j=0}^{\infty} \int_{\{a_n = n+j\}} P(x_1, \cdots, x_{n+j-1}; x_{n+j} = 0) - q\, dP \\
&= \sum_{j=2^{n-1}}^{2^n-1} \int_{\{a_n = j\}} P(x_1, \cdots, x_{j-1}; x_j = 0) - q\, dP \\
&= \int_{\{a_n = 2^n - 1\}} (p - q)\, dP + \cdots + \int_{\{a_n = 2^n - 1\}} (p - q)\, dP = p - q.
\end{aligned}
$$

6. **Asymptotic expectation theorems.** In this section we shall prove the theorems that stand in the same relation to the theorems of the preceding section as Theorem 8 stands to Theorem 5.

THEOREM 14. *If Ω is a real-valued stochastic process such that the functions $x_n(\omega)$ are uniformly bounded and such that $|E(x_1, \cdots, x_{n-1}; x_n)| < \epsilon$ for all $n \geqq n_0$ almost everywhere, then $|E(x_1', \cdots, x_{n-1}'; x_n')| < \epsilon$ for all $n \geqq n_0$ almost everywhere (where the x_n' are obtained from the x_n by a system transformation).*

Proof. Let I_{n-1}' be any measurable cylinder set over $x_1', \cdots, x_{n-1}'$. Then

$$
\begin{aligned}
\left| \int_{I_{n-1}'} E(x_1', \cdots, x_{n-1}'; x_n')\, dP \right| &= \left| \int_{I_{n-1}'} x_n'\, dP \right| \\
&= \left| \lim_{\delta \to 0} \sum_{r=-\infty}^{\infty} r\delta P(I_{n-1}'\{r\delta \leqq x_n' < (r + 1)\delta\}) \right| \qquad (26) \\
&= \left| \lim_{\delta \to 0} \sum_{r=-\infty}^{\infty} r\delta \sum_{j=0}^{\infty} P(I_{n-1}'\{a_n = n + j\}\{r\delta \leqq x_{n+j} < (r + 1)\delta\}) \right|.
\end{aligned}
$$

We now define the auxiliary function y to be x_{n+j} on $I_{n-1}'\{a_n = n + j\}$ and zero elsewhere. Then

$$
\begin{aligned}
&\left| \int_{I_{n-1}'} E(x_1', \cdots, x_{n-1}'; x_n')\, dP \right| \\
&\qquad = \left| \lim_{\delta \to 0} \sum_{r=-\infty}^{\infty} r\delta \sum_{j=0}^{\infty} P(I_{n-1}'\{a_n = n + j\}\{r\delta \leqq y < (r + 1)\delta\}) \right|
\end{aligned}
$$

$$= \left| \int_{\sum_{j=0}^{\infty} I'_{n-1}\{a_n=n+j\}} y\,dP \right| = \left| \sum_{j=0}^{\infty} \int_{I'_{n-1}\{a_n=n+j\}} x_{n+j}\,dP \right|$$

$$\leqq \sum_{j=0}^{\infty} \int_{I'_{n-1}\{a_n=n+j\}} |E(x_1, \cdots, x_{n+j-1}; x_{n+j})|\,dP$$

$$\leqq \sum_{j=0}^{\infty} \int_{I'_{n-1}\{a_n=n+j\}} \epsilon\,dP = \int_{I'_{n-1}} \epsilon\,dP \quad \text{for} \quad n \geqq n_0.$$

Since this is true uniformly in I'_{n-1}, we have $|E(x'_1, \cdots, x'_{n-1}; x'_n \epsilon E)| < \epsilon$, for all $n \geqq n_0$, almost everywhere. This concludes the proof of Theorem 14.

We note that the proof of this theorem remains unaltered if we remove the absolute value signs from its statement. This shows, for example, that if the conditional expectations are all negative, they will remain negative after any system transformation.

We note also that this theorem stands in the same relation to Theorem 8 as Theorem 10 to Theorem 5.

THEOREM 15. *If Ω is a real-valued stochastic process such that the functions $x_n(\omega)$ are uniformly bounded and such that $E(x_1, \cdots, x_{n-1}; x_n)$ converges to zero almost everywhere as $n \to \infty$, then $E(x'_1, \cdots, x'_{n-1}; x'_n)$ converges to zero in probability (where the x'_n are obtained from the x_n by a system transformation).*

Proof. A necessary and sufficient condition that $E(x'_1, \cdots, x'_{n-1}; x'_n)$ converge to zero in probability is that to every positive number ϵ there correspond a positive integer n_0, such that for $n > n_0$ $\left| \int_{\Lambda'} x'_n\,dP \right| < \epsilon$ whatever the measurable set Λ' over $x'_1, \cdots, x'_{n-1}$ may be. (By definition $\int_{\Lambda'} E(x'_1, \cdots, x'_{n-1}; x'_n)dP = \int_{\Lambda'} x'_n\,dP$.) We have now

$$\left| \int_{\Lambda'} x'_n\,dP \right| = \left| \lim_{\delta \to 0} \sum_{r=-\infty}^{\infty} r\delta P(\Lambda'\{r\delta \leqq x'_n < (r+1)\delta\}) \right| \tag{27}$$

$$= \left| \lim_{\delta \to 0} \sum_{r=-\infty}^{\infty} r\delta \sum_{j=0}^{\infty} P(\Lambda'\{a_n = n+j\}\{r\delta \leqq x_{n+j} < (r+1)\delta\}) \right|.$$

We define the function y to be x_{n+j} on $\Lambda'\{a_n = n+j\}$ and zero elsewhere. Then

$$\left| \int_{\Lambda'} x'_n\,dP \right| = \left| \lim_{\delta \to 0} \sum_{r=-\infty}^{\infty} r\delta \sum_{j=0}^{\infty} P(\Lambda'\{a_n = n+j\}\{r\delta \leqq y < (r+1)\delta\}) \right|$$

$$= \left| \int_{\sum_{j=0}^{\infty} \Lambda'\{a_n=n+j\}} y\,dP \right| = \left| \sum_{j=0}^{\infty} \int_{\Lambda'\{a_n=n+j\}} x_{n+j}\,dP \right| \tag{28}$$

$$\leqq \sum_{j=0}^{\infty} \int_{\Lambda'\{a_n=n+j\}} |E(x_1, \cdots, x_{n+j-1}; x_{n+j})|\,dP.$$

By hypothesis $E(x_1, \cdots, x_{n-1}; x_n)$ converges to zero almost everywhere. We apply Egoroff's theorem. Given any positive number ϵ, there exists a set D such that $P(CD) = 1 - P(D) < \epsilon$ and such that on D we have

$$|E(x_1, \cdots, x_{n+j-1}; x_{n+j})| < \epsilon \tag{29}$$

for n sufficiently large and $j = 0, 1, 2, \cdots$. Then

$$\begin{aligned} \left|\int_{\Lambda'} x'_n\, dP\right| &\leqq \sum_{j=0}^{\infty} \int_{\Lambda'\{a_n=n+j\}D} + \int_{\Lambda'\{a_n=n+j\}CD} |E(x_1, \cdots, x_{n+j-1}; x_{n+j})|\, dP \\ &\leqq \sum_{j=0}^{\infty} \int_{\Lambda'\{a_n=n+j\}} \epsilon\, dP + \int_{\Lambda'\{a_n=n+j\}CD} K\, dP, \end{aligned} \tag{30}$$

where K is the common upper bound of the functions x_n. (If $|x_n| < K$, $|E(x_1, \cdots, x_{n-1}; x_n)| \leqq K$.) Hence, finally,

$$\left|\int_{\Lambda'} x'_n\, dP\right| \leqq \epsilon P(\Lambda' D) + KP(\Lambda' \cdot CD) \leqq \epsilon + K\epsilon. \tag{31}$$

This concludes the proof of Theorem 15.

We note that the example of the preceding section shows that the convergence in probability of $E(x_1, \cdots, x_{n-1}; x_n)$ is not invariant under every system transformation.

Bibliography

I. Z. W. Birnbaum and J. Schreier, *Eine Bemerkung zum starken Gesetz der Grossen Zahlen*, Studia Math., vol. 4(1933), pp. 85–89.

II. A. H. Copeland, *Admissible numbers in the theory of geometrical probability*, Am. Jour. of Math., vol. 53(1931), pp. 153–162.

III. A. H. Copeland, *Admissible numbers in the theory of probability*, Am. Jour. of Math., vol. 50(1928), pp. 535–552.

IV. J. L. Doob, *Note on probability*, Annals of Math., vol. 37(1936), pp. 363–367.

V. J. L. Doob, *Probability and statistics*, Trans. Amer. Math. Soc., vol. 36(1934), pp. 759–775.

VI. J. L. Doob, *Stochastic processes with an integral-valued parameter*, Trans. Amer. Math. Soc., vol. 44(1938), pp. 87–150.

VII. H. Hahn, *Über die Multiplikation total-additiver Mengenfunktionen*, Annali della R. Sc. Norm. Sup. di Pisa, vol. 2(1933), pp. 429–452.

VIII. E. Hopf, *On causality, statistics, and probability*, Jour. of Math. and Phys., vol. 13 (1934), pp. 51–102.

IX. H. Huntemann, *Über den mathematischen Kern des Prinzips vom ausgeschlossenen Spielsystem und eine darauf gegründete Wahrscheinlichkeitstheorie*, Deutsche Mathematik, vol. 2(1937), pp. 593–622.

X. A. Kolmogoroff, *Grundbegriffe der Wahrscheinlichkeitsrechnung*, Berlin, 1933.

XI. R. von Mises, *Wahrscheinlichkeitsrechnung*, Wien, 1931.

XII. S. Saks, *Theory of the Integral*, Warsaw, 1937.

XIII. A. Wald, *Die Widerspruchfreiheit der Kollektivbegriffes der Wahrscheinlichkeitsrechnung*, Ergebnisse eines math. Kolloq., vol. 8(1936), pp. 38–72.

University of Illinois.

Ve
D

K,

nce
em

sen

Am.

. of

34),

ner.

ells

1. 13

enen
sche

3.

eits-

Ar

Cl

Co

Du

Ec

Hi

Th

La

Th

Th

PERIODICALS PUBLISHED BY DUKE UNIVERSITY

American Literature. A quarterly journal devoted to research in American Literature, published with the coöperation of the American Literature Group of the Modern Language Association of America. Subscription, $4.00 per year. Back volumes, $5.00 each.

Character and Personality. A psychological quarterly devoted to studies of behavior and personality. Subscription, $2.00 per year. The first number was published September, 1932.

Contributions to Psychological Theory. A monograph series dealing with problems of psychological theory in the widest sense, including their relations to other fields of inquiry. The monographs appear irregularly. Subscription, $5.00 per volume of approximately 450 pages.

Duke Mathematical Journal.

Ecological Monographs. A quarterly journal devoted to the publication of original researches of ecological interest from the entire field of biological science. Subscription, $6.00 per year. The first number was published January, 1931.

Hispanic American Historical Review. A quarterly review dealing with the history of the Latin-American countries. Subscription, $4.00 per year.

The Journal of Parapsychology. A scientific quarterly dealing with telepathy, clairvoyance, and other parapsychological problems. Subscription, $3.00 per year.

Law and Contemporary Problems. A quarterly published by the School of Law, presenting in each issue a symposium on a problem of current importance having significant legal aspects. Subscription, $2.00 per year. The first number was published September, 1933.

The South Atlantic Quarterly. A magazine of modern opinion and discussion, founded in 1902. Subscription, $3.00 per year.

The Southern Association Quarterly. As official organ of the Southern Association of Colleges and Secondary Schools, it contains the proceedings of the annual meeting, together with much additional material directly related to the work of the Association. Subscription, $4.00 per year.

DUKE UNIVERSITY PRESS
DURHAM, NORTH CAROLINA

CONTENTS